Equilibrium Sets of Some $GI/M/1$ Queues (with more examples)

N. Hemachandra Sandhya Tripathi Kishor Patil

October 18, 2016

Abstract

We view queue as a service facility and consider the situation when users offer arrival rate at stationarity that depends on the Quality of Service (QoS) they experience. We study the equilibrium points and the equilibrium sets associated with this interaction and their interpretations in terms of business cycles. The queue implements the optimal admission control (either discounted or ergodic) policy in the presence of holding cost and admission charge for admitted customers; threshold policies are known to be optimal for such queues. We consider two QoS measures: the long run fraction of customers lost, L, and the long run rate of customers lost, L_1. Our first result is that both the QoS measures for $GI/M/1$ queues are locally continuous with respect to the arrival rate. For $M/M/1$ queues we show that control limit is finite and that (Assumption A1) the QoS measures increase with arrival rate. We also show that multiple optimal policies lead to equilibrium sets. We observe that various combinations of cost criteria and QoS measures lead to differing equilibrium behaviour. Under A1, similar results hold for $D/M/1$ queue. We study a $GI/M/1$ queue with discrete arrival rate supports whose arrival rate is locally continuous. We illustrate that A1 need not hold, but a relaxed assumption may hold. Nonetheless, change of support is another cause for emergence of equilibrium sets and equilibrium behaviour is interesting as there may be multiple equilibrium points/sets, etc. Generalized equilibrium sets may also exist; these are usually due to the non-contraction nature of the QoS measure, the rate of customers lost, L_1. We present details of some computational examples that were are not presented in (8).

keywords: admission control of queues; quality of service; multiple optimal policies; change of support; invariant sets; generalized equilibrium sets; non-contraction maps; fixed points; parametrized MDPs

1 Introduction

We view a queue as a service facility (service-provider) and consider the situation when customer base (user-set) offers an arrival rate at stationarity that depends on the Quality of Service (QoS) they experience. In our model, the queue uses a revenue optimal policy and the QoS experienced by the user-set is induced by this revenue optimal policy. We are primarily interested in the equilibrium points and the equilibrium sets associated with this interaction between the queue and user-set. Many present day technological systems like communication networks, transportation systems, etc., that offer various services have congestion which can be captured by queueing models; apart from the price of the service, users are also concerned with the service levels they obtain while using these service systems.

To analyse the equilibrium behaviour, we consider the relation between the QoS offered by the queue and the user-set offered arrival rate λ in the form of firm-market interaction function $f(\cdot)$ as defined in (9). The arrival rate λ offered by the user-set when it experiences a QoS at *stationarity* is given by $f(\text{QoS})$. In fact, we assume $f(\cdot)$ to be as below (9),

$$f(L(\pi^*_\lambda, P(\lambda))) = m - eL(\pi^*_\lambda, P(\lambda)) \tag{1}$$

where m and e are given constants so that the arrival rate, $\lambda \in (0, m)$, $L(\pi^*_\lambda, P(\lambda))$ is the stationary QoS measure felt by the user-set due to the use of the revenue optimal policy $\pi^*_\lambda (=: \pi^*(\lambda))$ deployed by the service-provider (queue) and $P(\lambda)$ is the price charged by the service facility queue; both π^*_λ and $P(\lambda)$ can depend on the user-set offered arrival rate λ. In $m - eL(\pi^*_\lambda, P(\lambda))$, known as the affine (linear) demand model, m is the maximal arrival the user-set can offer (market capacity) and e is the arrival rate elasticity to the QoS $L(\cdot, \cdot)$. Figure 1 gives an overview of the problem we address; (9) considers a discrete time model where λ is the mean demand in each period, but, in below we stick to the queueing convention where λ is called the arrival rate. The basic result is that under mild conditions, either an equilibrium point or an equilibrium set (an invariant set, not a set of equilibrium points, details are in Section 2) exists. Existence of an equilibrium set offers one possible explanation for the business cycles between high arrival rate regime and low arrival rate regime that we observe in the operation of many service systems; after a large enough time, there is a regime switch in the user-set offered arrival rate $f(\cdot, \cdot)$.

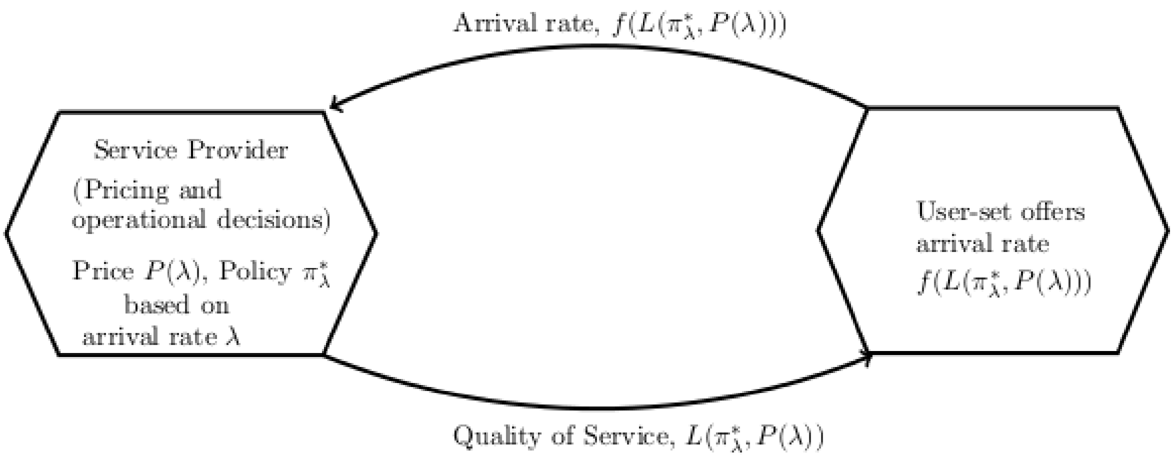


Figure 1: A schematic for service-provider (queue) user-set (customer base) interaction (9). In rest of the paper, we denote the optimal policy (π^*_λ) by R^*_λ.

In (9) it was observed that the difference between the service-provider user-set interaction in discounted and average reward Markov decision process (MDP) model for vehicle relocation systems, a service facility prevalent on some educational campuses or small towns, is negligible and the equilibrium points and sets are nearly the same. It was pointed out that queuing systems with holding costs (in addition to admission revenue) should provide an example of the service-provider user-set interaction MDP model where this difference in equilibrium is non-trivial. In addition to the long run fraction of customers lost $L(\lambda, \pi^*_\lambda)$, we also consider another user-set time average QoS measure $L_1(\lambda, \pi^*_\lambda)$, the rate of customers lost. We identify various aspects like cost criteria, etc., that effect this interaction. We note that in stylized models of $M/M/1$ and $D/M/1$ queues, the QoS measures (and hence $f(\lambda)$ from (1)) are monotone wrt arrival rate λ (assumption A1) and show that the presence of multiple revenue optimal policies is the cause for the emergence of equilibrium sets. For a certain $GI/M/1$ queue a change in support of inter-arrival times is another cause. Additionally, we note that this interaction can be quite complicated; for such queues, there can be more than one equilibrium set or point or the iterates of $f(\cdot,\cdot)$ need not converge to an equilibrium point as they now lie in an equilibrium set, etc. This also means that this interaction can experience multiple regimes of rate like medium and high arrival rates with various cycle lengths; this leads us to the notion of generalized equilibrium sets wherein a relaxed version of A1 holds. The choice of the QoS measures by the user-set can also cause these equilibrium sets. There can be discontinuities in the optimal value of the queue as a function of arrival rate, say, because of multiple optima or change in the support of the arrival distribution. The choice of these optimal policies, particularly at these discontinuities, means that QoS measures depend on the way they are defined at these discontinuities which in turn influences the nature of the equilibrium points and sets.

Pinhas Noar (10) pioneered the analysis of the admission control of queues. Stidham in a series of papers (18), (16), (15), etc., established various structural results for these models, including the insight that individual optimal admission rate is no less than that of the social optimal one; interpreting the difference to the externality that each admitted customer imposes on others in form of the increased congestion they experience and emphasizing the role of admission fee in admission control models. We extensively rely on the results and efficient algorithms of Van Nunen and Puterman (3) and Puterman and Thomas (6); we mention some minor typos. The book by Puterman (5) is a comprehensive book for Markov decision models and in particular, Chapter 11 treats semi-Markov control models including the admission control of $GI/M/1$ queues. The monograph by Sennott (4) treats numerous types of control problems in queues, including admission control problems. We use the monotonicity results of Cil, et al. (1) and Rue and Rosenshine (14).

Kim (21) proposed an optimization model for computing optimum revenue using static pricing in a two class (premium (circuit-switched) class and best effort (packet-switched) class) queueing model with QoS-based-demand and demand-based-QoS fixed point constraints. Convergence results of the proposed numerical schemes for computing these fixed points are presented. We are more interested in the existence of equilibrium sets and points. Many game theoretic aspects in queueing are studied by Hassin and Haviv (12) and comprehensively summarised by Hassin (11).

Stidham (17) considers many models involving static optimization (open loop control) where each customer has a utility as well as a congestion cost that depend on λ and the decision maker chooses an arrival rate that maximizes the difference between utility and congestion cost. These models broadly fall into one of these three types; the decision maker can be an individual customer (individual optimality) or considers the aggregate/collective utility of all customers (social optimality) or queue (facility) manager who considers the net profit per unit time. Many variants like networks of queues, etc., are also nicely analysed in the monograph (17). The optimal arrival rates can be interpreted as arrival rates at a certain equilibrium that is different from ours. Because, in our setting, the decision maker (queue manager) uses a dynamic control (closed loop control) for any given λ that maximizes either the discounted or ergodic profit that trades off the admission fee collected from the admitted customers against the holding/congestion costs incurred by it; the collective of users (user-set), while offering the demand, is interested in the QoS it experiences induced by the revenue optimal policy of the queue manager. Also, it turns out that utility and congestion functions considered in (17) are usually differentiable convex/concave functions of λ; in our case, the key function, $f(\lambda)$ is discontinuous which can lead to equilibrium sets.

In Section 2, we present basic results like existence of equilibrium sets. Section 3 gives details including causes of equilibrium sets, QoS computations, numerous examples, generalizations of above, etc. We close with some discussion in Section 4. Most proofs are deferred to Appendices.

2 Equilibrium sets in queues: Existence and properties

In this section we define the generic equilibrium sets. We investigate their existence in the above setting and present an interpretation.

2.1 Existence of equilibrium sets

The queue (service-provider) in our setting uses a revenue optimal policy R^*_λ of admission control as depicted in Figure 2. Inter-arrival times of jobs are i.i.d. (with a given distribution having arrival rate λ) and a single server is available to serve the jobs at an given exponential rate μ. There is a control on admitting the new arrival/job to the queue. It can accept the job or may reject it depending on state of the queue, the number in the queue at that epoch, including the one being served. An entering customer pays the queue a service charge of r for the utility of the availed service. A holding cost h per customer per unit time is incurred to the queue. The service-provider (queue) is interested either in the optimal discounted or ergodic revenue policy corresponding to these costs and rewards stochastic processes over the infinite horizon.

The above continuous time semi-Markov optimal (either ergodic or discounted) control problem can be cast as a discrete time Markov decision model by a suitable embedding, see for example Puterman (5). Stidham (18) showed that if holding cost h is a convex function of the number of customers in the system, control limit policies, i.e., policies of threshold type are revenue optimal for both discounted and average cost criteria. The queue determines revenue optimal discounted threshold $R^*_{\lambda,dc}$ or ergodic threshold $R^*_{\lambda,av}$ when the arrival rate is λ; these induce QoS measure $L(\lambda, R^*_{\lambda,dc})$ or $L(\lambda, R^*_{\lambda,av})$ respectively. To ease the notation, we will just write this revenue optimal threshold as R^*_λ; the cost criteria will be clear from the context. Also, we abbreviate QoS measures as

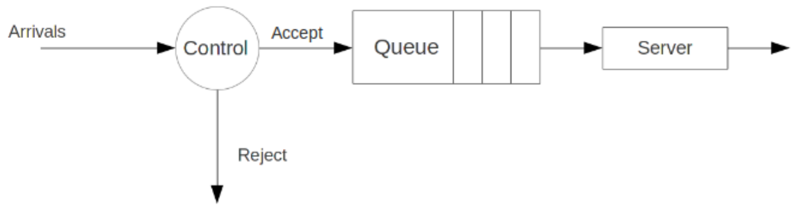


Figure 2: Admission control queueing system

$L(\lambda)$. Performance measures like mean waiting time or customers lost, etc., are potential candidates for measuring the quality of service being offered. We work with two QoS measures: (long run) fraction of customers lost $L(\lambda)$ and (asymptotic) rate of customers lost $L_1(\lambda)$ (more details are in Section 3). The user-set (market) offers arrival rate $f(QoS)$ based on the QoS it experiences, for a given $f(\cdot)$. We are interested in the equilibrium point or equilibrium sets in this interaction. The firm-market interaction function $f(\cdot)$ in (1), that captures this interaction, should satisfy the following assumption.

Assumption 1 (A1): $f(\lambda, L(\lambda, R^*_\lambda))$ is a non increasing function of λ; for simplicity we denote $f(\lambda, L(\lambda, R^*_\lambda))$ by $f(\lambda)$.

Definition 1 (Equilibrium Set (9)). *If there exists some $l_0 \in (0, m)$ and intervals $R := (f(l_0), \lim_{\lambda \uparrow l_0} f(\lambda))$, $E_1 := (0, f(l_0)]$ & $E_2 := [\lim_{\lambda \uparrow l_0} f(\lambda), m)$ and $E := E_1 \cup E_2$ such that*

- *for $\lambda \in R$, $f(\lambda) \in E$*
- *for $\lambda \in E$, $f(\lambda) \in E$ ($\lambda \in E_1$ implies $f(\lambda) \in E_2$ and $\lambda \in E_2$ implies $f(\lambda) \in E_1$),*

then E and R are called equilibrium set and repelling set of service-provider user-set interaction respectively.

Remark 1. *The equilibrium set is a generalization of the equilibrium point and is an invariant set under $f(\cdot)$. Under assumption A1 and other conditions below, (9) shows existence of an equilibrium set when an equilibrium point doesn't exist in the above framework.*

To show the existence of equilibrium sets or points in the above queue and user-set interaction as in (9), we need that the underlying MDP has finite state space and finite action space, piecewise continuity of the QoS measures L and L_1 and that assumption $A1$ holds.

Van Nunen and Puterman (3) observe that when the discount parameter α is less than h/r for $M/M/1$ and $D/M/1$ queues, then the optimal control limit, R^*_λ is finite. Throughout this paper we assume this condition for these queues. The following theorem shows that the optimal control limit for average reward $M/M/1$ is finite; we defer the proof to Appendix A.1.

Theorem 1. *For an average reward stable $M/M/1$ queue, there always exists a finite admission control limit R^*_λ.*

For an admission controlled $GI/M/1$ queue with countable state space, when optimal control limit R_λ^* is finite, consider the related admission control problem in which state space is truncated as $(0,\ 1,\ 2,\ \ldots\ R_\lambda^*,\ R_\lambda^*+k)$, with $k \geq 1$. Optimal control limit for this problem is still R_λ^*. Since we always consider situations wherein the optimal control limit is finite, the admission controlled queuing MDP can be taken as a finite state-action MDP with the same optimal policy R_λ^*.

One QoS measure we consider is

$$L(\lambda) = \text{long run fraction of customers lost when the arrival rate is } \lambda. \tag{2}$$

The other QoS measure which we consider is the asymptotic rate of customers lost,

$$L_1(\lambda) = \lambda L(\lambda). \tag{3}$$

The average cost semi-Markov admission control problem of a $GI/M/1$ can be viewed as a discrete time MDP which in turn can be viewed as a suitable Linear Program (LP), (5). If data of this control problem like transition probabilities are continuous functions of rate λ, relying on the continuity of the sets of optimizers of these LPs, Theorem 1 of (9) says that the optimal revenue value at state s, $g^{R_\lambda}(s) := g^R(s,\lambda)$ is continuous over an interval (u,v) about λ when the optimal policy R_λ^* at λ is unique. We assume that arrival distributions are such that the optimal revenue is right continuous at end point u of the interval (u,v). Then, it follows that time averages associated with the revenue optimal policies are also piecewise continuous functions of λ. When the revenue optimal policy, R_λ at a λ is not unique, say $\{R_{1,\lambda}^*,\cdots,R_{l,\lambda}^*\}$ for some $l \geq 2$ are revenue optimal policies, we have that $g_{R_\lambda^*}(s) = \lim_{\lambda_n \downarrow \lambda} g_{R_{\lambda_n}^*}(s)$ for some $R_\lambda^* \in \{R_{1,\lambda}^*,\cdots,R_{l,\lambda}^*\}$. Then, for a given time average, we define it at λ as per this policy R_λ^* so that the time average is also right continuous at λ; see (9) for details. The time average values at the end points of the interval (u,v) are similarly defined.

Since the QoS measure, fraction of customers lost, L can be seen as a ratio of two time averages (long run rate of customers lost and long run rate of arrivals), we can invoke Theorem 2 of (9), and claim that L is piecewise continuous over $(0,m)$. It then follows that L_1 is also a piecewise continuous function over $(0,m)$. Such results hold in the discounted revenue model also.

In case of admission controlled $M/M/1$ queue with discounted reward criterion, Cil et. al. (1) have showed that optimal policy, i.e., control limit R_λ^* is decreasing in arrival rate λ and hence, $f(\lambda, L(\lambda, R_\lambda^*))$ is a non increasing function of λ. Rue and Rosenshine (14) considered an admission controlled $M/M/1$ queue where customer is the decision maker, i.e., customer takes the decision of entering or not entering the queue, to maximize the sum of individual net benefits called socially optimal criteria. In this case, they showed that the average reward criterion based optimal policy (threshold type) is a non-increasing function of λ. It can be easily seen that the objective which is maximized in either customer as a decision maker or queue manager as decision maker, is identical. Therefore, in case of admission controlled $M/M/1$ queue with both discounted and average reward criterion, assumption $A1$ holds.

We are considering four models, depending on the revenue criteria (discounted or ergodic) of the queue and the QoS measure ($L(\cdot)$ or $L_1(\cdot)$) of the user-set.

In view of the above three observations for each of these four models, we then have,

Theorem 2. *For an admission controlled $M/M/1$ queue, when queue uses ergodic revenue criteria and user-set uses QoS $L(\cdot)$, either an equilibrium point or the equilibrium set will exist. Either an equilibrium point or an equilibrium set also exist in the other* 3 *models.*

We will see in Section 3 that one reason for emergence of equilibrium sets in $M/M/1$ queue is the presence of multiple revenue optimal policies. The support of inter-arrival times for these queues is the whole of R_+. We now consider $D/M/1$ queues; the support of inter-arrival times for these queues is just a singleton. We assume that for $D/M/1$ queue, $A1$ holds; examples validating this assumption will be provided in Section 3.6. As shown above the admission control problem is a finite state and finite action MDP and QoS measures are locally continuous. Thus, one has,

Theorem 3. *For an admission controlled $D/M/1$ queue that uses average cost criteria and user-set that uses QoS $L(\cdot)$, under the assumption $A1$, either an equilibrium point or an equilibrium set will exist. Under $A1$, either an equilibrium point or an equilibrium set will also exist in the other* 3 *models.*

Next, we are interested in the effect of the change in the support of the arrival distribution on the equilibrium behaviour. To understand this, we consider queues where arrivals per unit time follow a particular form of discrete distribution whose mean still is *locally continuous*. In Section 3.7, we will show that $A1$ need not hold for such queues and see the implications; the other aspects of the relevant MDP and QoS measures for these queses hold.

2.2 Properties of equilibrium sets

An equilibrium point in above will exist if there is a fixed point for the basic firm-market interaction equation $\lambda = f(\lambda, \cdot)$. If this function has discontinuity, then there is a possibility that the fixed point may not exist and in view of above, there will be an equilibrium set.

This discontinuity in QoS can be directly attributed to multiple optimal control limits or to the change in support of the inter-arrival times as in the case of discrete arrival distribution. Two other indirect causes for emergence of equilibrium sets are the changes in the cost criteria of the queue manager or the change in the QoS measure by the user-set (market). All these four causes can be grouped into two categories as follows:

1. Queue's (firm) role:
   - Multiple control limits
   - Ergodic or discounted criteria
2. User-set's (market) role:
   - Support of the arrival distribution
   - Type of the QoS measure

The details of these causes along with the examples is given in Section 3.4, 3.5, 3.6 and 3.7.

For a $M/M/1$ queue with the queue manager using the discounted cost structure and user-set using L as a QoS measure, Figure 3 gives an example of sets E_1 and E_2 and the equilibrium set E. In (9) the following interpretation of equilibrium sets was proposed. Suppose market offers some high arrival rate, say $\lambda_0 \in E_2$ and user-set is interested in one of the QoS measures, say $L(\cdot)$. Then poor service is offered (due to system capacity limitations at this high arrival rate) and after long enough time when the system is nearly in steady state (so that QoS experienced by user-set is approximately $L(\cdot)$), the user-set responds by offering arrival rate $f(\lambda_0) =: \lambda_1$, which is a lesser arrival rate as $\lambda_1 \in E_1$. Since arrival rate is decreased, again in long enough time the system will start working smoothly and arrival rate offered by the user-set $f(\lambda_1) =: \lambda_2$ increases and is such that $f(\lambda_1) \in E_2$. However, when the arrival rate increases, system is not able to handle it properly leading to discontent among the customers and again $\lambda_3 := f(\lambda_2) \in E_1$. And, this process continues. Thus, the queue (firm) is switching between a high arrival rate regime and a low arrival rate regime. If an equilibrium set is present, the resultant toggling between E_1 and E_2 is one of the explanations of the business cycles in QoS based service facilities like queues experience. The toggling between E_1 and E_2 is also illustrated in Figure 3.

The set $R = (0, m) \setminus (E_1 \cup E_2)$ is the repelling set; for all $\lambda \in R$, we have $f(\lambda) \in E$. If the arrival rate offered by the user-set is from R, after a reasonable amount of time when the queue is nearly in steady state, the user-set offers an arrival rate from one of the E_i's of the equilibrium set E; this then leads to above associated oscillations as illustrated in Figure 3.

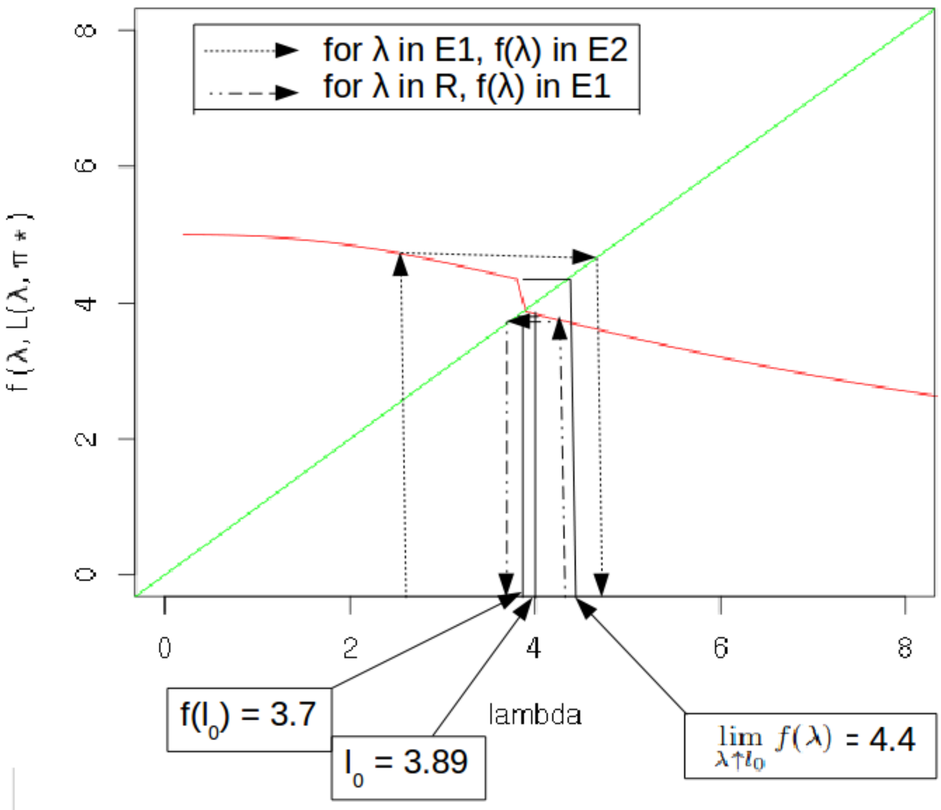


Figure 3: For a $M/M/1$ queue with costs and parameters $h = 1$, $r = 0.4$, $\mu = 5.5$, $\alpha = 0.54$, $m = 5$ and $e = 4$ using discounted cost criteria and QoS measure L, we get $l_0 = 3.89$, $f(l_0) = 3.7$ and $\lim_{\lambda \uparrow l_0} f(\lambda) = 4.4$. Hence, the equilibrium set is $E = E_1 \cup E_2$ where $E_1 = (0, 3.7]$ and $E_2 = [4.4, 5)$ and repelling set is $R = (3.7, 4.4)$.

Another representation of equilibrium sets can be given by iteratively computing the sequence $\{\lambda_i\}_{i\geq 0}$ as

$$\lambda_{i+1} = f(\lambda_i), \quad \text{with} \quad \lambda_0 \in (0, m)$$

If these $\{\lambda_i\}_{i\geq 0}$ iterates converge to a point, then we conclude that an equilibrium point exists, otherwise an equilibrium set exists. We illustrate the existence of equilibrium sets or equilibrium point in most of the examples in terms of these iterates $\{\lambda_i\}_{i\geq 0}$.

3 Equilibrium sets for $M/M/1$ queue, $D/M/1$ queue, and queues with discrete arrival rate distribution

In this section, we give characterization of QoS measures L and L_1, investigate various causes of equilibrium sets, including the role of multiple revenue optimal policies and change of support of arrival rate, and present numerous examples illustrating various types of equilibrium sets. Before that we briefly present the relevant MDP formulation of the admission control of $GI/M/1$ queue (discounted and average cost criteria cases) (5).

3.1 Discounted cost model

Let $\{T_n\}$ be a sequence of iid random variables representing time between the arrival of n^{th} and $(n-1)^{th}$ customer to the queue, with distribution function $F(\cdot)$. Also, $N(t)$ be the number of service completions in a period of length t. Now, if α is the rate at which rewards and costs are discounted continuously, then the expected discounted holding cost, $c(s)$ can be obtained as follows (3):

$$c(s) = E\left\{\int_0^{T_1} e^{-\alpha t} h((s - N(t))^+) dt\right\}$$

If $\{x_n^\pi; n \geq 0\}$ is the embedded Markov chain of the semi-Markov process recording the number in the queue under the policy $\pi = (\delta_0, \delta_1, \ldots)$, then the infinite horizon expected reward of this policy is given by

$$v_\alpha^\pi(s) = E_s^\pi\left\{\sum_{n=0}^{\infty} exp\left(-\alpha \sum_{i=1}^{n} T_i\right)[r\delta_n(x_n^\pi) - c(x_n^\pi + \delta_n(x_n^\pi))]\right\} \tag{4}$$

where E_s^π is the expected value with respect to the process determined by π and $x_0^s = s$. Using optimal policy of control limit type, the infinite horizon expected reward is,

$$v_\alpha^{R_\lambda}(s) = \begin{cases} r - c(s+1) + \sum\limits_{j=0}^{s} P_j v_\alpha^{R_\lambda}(s+1-j) + q_{s+1} v_\alpha^{R_\lambda}(0), & s < R_\lambda \\ -c(s) + \sum\limits_{j=0}^{s-1} P_j v_\alpha^{R_\lambda}(s-j) + q_s v_\alpha^{R\lambda}(0), & s \geq R_\lambda. \end{cases} \tag{5}$$

where, $P_j = \int\limits_0^\infty e^{-\alpha t} \frac{e^{-\mu t}(\mu t)^j}{j!} dF(t)$ and $q_j = \sum\limits_{k=j}^{\infty} P_k$.

Admission control in $GI/M/1$ queues is explored in detail by Van Nunen and Puterman (3). They exploited the special structure of such a queue to develop an efficient algorithm which calculates optimal control limit, R_λ^* and we use it. However, we observe some minor typos in expressions of a_{R_λ} and b_{R_λ} given in (3). The correct expressions are as follows,

$$a_{R_\lambda} = P_0^{-1}\Bigg[1 - P_1 - \sum_{k=1}^{R_\lambda - 1} \frac{P_{k+1}}{\prod_{i=1}^{k} a_{R_\lambda - i}} - \frac{q_{R_\lambda + 1}}{\prod_{i=1}^{R_\lambda} a_{i-1}}\Bigg]$$

$$b_{R_\lambda} = P_0^{-1}\Bigg[r - c(R_\lambda + 1) + \sum_{j=2}^{R_\lambda} P_j \Big[\sum_{k=2}^{j} \frac{b_{R_\lambda + 1 - k}}{\prod_{i=k}^{j} a_{R_\lambda + 1 - j}}\Big] + q_{R_\lambda + 1}\Big[\sum_{j=0}^{R_\lambda - 1} \frac{b_j}{\prod_{i=0}^{j} a_i}\Big]\Bigg]$$

3.2 Average cost model

To calculate the control limit for average case, we consider the model described in Chapter 11 of (5). Let $v_\alpha^{R_\lambda}$ be the expected infinite horizon total discounted reward and $v_t^{R_\lambda}(s)$ be the expected total reward up to time t starting in state s, when the policy is R_λ. Then, we define expected long run average reward g^{R_λ} for this policy as,

$$g^{R_\lambda} = \lim_{\alpha \to 0} \alpha v_\alpha^{R^\lambda}(s) = \lim_{t \to \infty} \frac{1}{t} v_t^{R_\lambda}(s)$$

This g^{R_λ} can be found by solving Equation 13 in (6). However, limits of summation should be changed and one more equation is needed in addition to system of equation as,

$$w_{R_\lambda}(s) = \begin{cases} r - c(s+1) - g^{R_\lambda} y + \sum_{j=0}^{s} P_j w_{R_\lambda}(s+1-j) + q_{s+1} w_{R_\lambda}(0) & 0 \le s \le R_{\lambda - 1} \\ -c(s) - g^{R_\lambda} y + \sum_{j=0}^{s-1} P_j w_{R_\lambda}(s-j) + q_s w_{R_\lambda}(0) & s = R_\lambda. \end{cases} \tag{6}$$

where y denotes the expected length of time until the next decision epoch i.e. inter-arrival time. This is same as Equation 11.4.13 in (5).

In line with the model, $w_{R_\lambda}(0) = 0$ and $c(s)$ and P_j are as defined by Van Nunen and Puterman (3) with $\alpha = 0$. Theorem 2 in (6) states that if R_λ^* is the largest average optimal control limit, g^{R_λ} is monotone non-decreasing in R_λ for $R_\lambda < R_\lambda^*$ and monotone non-increasing in R_λ for $R_\lambda > R_\lambda^*$. We solve Equation 6 for each R_λ and plot g^{R_λ}. The optimal control limit R_λ^* is the control limit for which g^{R_λ} achieves maximum value. This method has been used to compute the control limit for all 3 types of queues considered in this paper.

Remark 2. *We showed earlier that the state space can be considered as finite and hence Blackwell Optimality (5) holds. Thus, the optimal control limit for average reward can be obtained via Van Nunen and Puterman algorithm by using very low discounting factor. However, finding the low enough discounting factor for which Blackwell optimality holds can be a cumbersome process.*

3.3 QoS measures for $GI/M/1$ queues

In this section, we define and give details for computing the quality of service measures for $GI/M/1$ queues under admission control.

3.3.1 QoS measures for $M/M/1$ queues

In case of Poisson arrivals, the stationary probability that system has n customers when R^*_λ is the control limit is given as follows :

$$\pi_{R^*_\lambda}(n) = \begin{cases} \frac{(1-\rho)\rho^n}{1-\rho^{n+1}} & \rho \neq 1 \\ \frac{1}{n+1} & \rho = 1 \end{cases} \qquad n = 0, 1, 2, \ldots, R^*_\lambda \tag{7}$$

Since the inter-arrival time is exponential, we use Poisson Arrivals See Time Averages (PASTA) (13) to calculate the probability that arriving customer sees the R^*_λ customers in the system. Thus, the two QoS measures can be defined as follows:

- Fraction of customers lost $L(\lambda, R^*_\lambda) = \begin{cases} \frac{(1-\rho)\rho^{R^*_\lambda}}{(1-\rho^{R^*_\lambda+1})} & \text{if } \rho \neq 1 \\ \frac{1}{R^*_\lambda+1} & \text{if } \rho = 1 \end{cases}$.
- Rate of customers lost $L_1(\lambda, R^*_\lambda) = \begin{cases} \frac{\lambda(1-\rho)\rho^{R^*_\lambda}}{(1-\rho^{R^*_\lambda+1})} & \text{if } \rho \neq 1 \\ \frac{\lambda}{R^*_\lambda+1} & \text{if } \rho = 1 \end{cases}$.

3.3.2 QoS measures for queues with deterministic and discrete support of the arrival rate distribution

The average expected reward g^{R_λ} can be used to derive a system of linear equations whose solution is the QoS for queues with non-Poisson arrivals and exponential service times as given in the proof in Appendix A.2.

Theorem 4. *If $\tilde{g}^{R_\lambda}$ is the solution of following system of equations,*

$$w_{R_\lambda}(s) = \begin{cases} 1 - g^{R_\lambda} y + \sum_{j=0}^{s} P_j w_{R_\lambda}(s+1-j) + q_{s+1} w_{R_\lambda}(0) & 0 \leq s \leq R_\lambda - 1 \\ -g^{R_\lambda} y + \sum_{j=0}^{s-1} P_j w_{R_\lambda}(s-j) + q_s w_{R_\lambda}(0) & s = R_\lambda. \end{cases} \tag{8}$$

where R_λ is control limit and y is expected inter-arrival time, then the QoS measures experienced by the user-set are,

$$L(\lambda, R_\lambda) = (1 - \tilde{g}^{R_\lambda} y) \quad \textit{and} \quad L_1(\lambda, R_\lambda) = \lambda \times (1 - \tilde{g}^{R_\lambda} y)$$

3.4 Multiple optimal control limits in admission controlled $GI/M/1$ queues

When the optimal control limit changes, there is a discontinuity in QoS measures of $GI/M/1$ queue. Due to the piecewise continuous nature of QoS, there could be possibly multiple optimal solutions (control limits) at the points of discontinuity of $f(\cdot)$.

From the nature of the solution of the system of equations in (5), we have this result whose proof is in Appendix B.1,

Theorem 5. *Assuming that there exists a unique optimal discounted control limit at λ_0 and the value vector $v_\alpha^{R_\lambda}$ is continuous at λ_0, there exists an open interval around λ_0 and the optimal control limit for all arrival rates in this interval is unique and is same as that of at λ_0.*

Now, we show the existence of multiple optimal control limits for an admission controlled $GI/M/1$ queue under discounted cost criteria; we defer the proof to Appendix B.2.

Theorem 6. *Given two intervals (λ_1, λ_2) and (λ_2, λ_3) with unique optimal discounted control limits R_λ^* and $\tilde{R}_\lambda^*$ and that the value vector $v_\alpha^{R_\lambda}$ is continuous at λ_2, multiple optimal control limits at λ_2 are R_λ^* and $\tilde{R}_\lambda^*$ (possibly along with others).*

Remark 3. *The above results hold for average cost criteria as the proof of Theorem 5 and 6 can be replicated for average cost criteria by replacing $v_\alpha^{R_\lambda^*}$ with $g^{R_\lambda^*}$; assuming $g^{R_\lambda^*}$ is a continuous function of λ.*

Remark 4. *In case of queues with discrete support of arrival rates, there could be discontinuity in the value vector $v_\alpha^{R_\lambda}$ and g^{R_λ} when the support changes. Hence, Theorem 5 and 6 will not be valid for the arrival rate ($\tilde{\lambda}_2$) at end point of the intervals ($(\tilde{\lambda}_1, \tilde{\lambda}_2]$ and $(\tilde{\lambda}_2, \tilde{\lambda}_3]$) generated using two different supports ($\{1, k, k+1\}$ and $\{1, k+1, k+2\}$); more details in Section 3.7 and illustration in Example 8.*

Remark 5. *Puterman and Thomas (6) showed in Corollary 3 that for a admission controlled $GI/M/1$ queue using discounted cost criteria, the multiple optimal control limits necessarily occur consecutively. Haviv and Puterman (7) showed that for $M/M/1$ queue under average cost criteria model, at a point λ there can exist only two multiple optimal control limits and they are consecutive integers R_λ^* and $R_\lambda^* + 1$; this leads to the theme of sensitive optimality criteria, (5). On the other hand, the above relates the optimal control limits over adjacent intervals to the multiple optimal control limit at the common point of these intervals.*

An interpretation of multiple optimal control limits can also be given in terms of value vectors. The control limit R_λ and $R_\lambda - 1$ can be optimal simultaneously if admitting one more customer is not changing the value vector. This observation leads to the following result whose proof is given in Appendix B.3; all components of the vectors need not be compared.

Theorem 7. *For an admission controlled $GI/M/1$ queue under discounted cost criteria, the condition*

$$v_\alpha^{R_\lambda}(0) = v_\alpha^{R_\lambda+1}(0)$$

is sufficient for checking the existence of multiple optimal control limits.

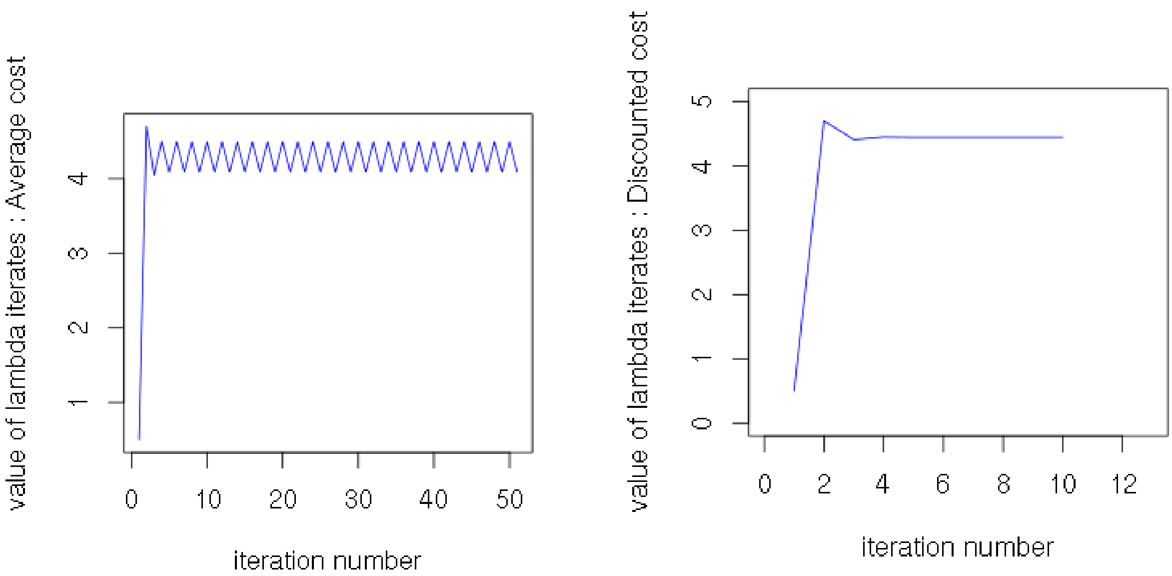
Using closed form expression for $v_\alpha^{R_\lambda}(0)$ in case of $M/M/1$ queue, Theorem 7 can be used to develop methods for calculating λ when $R_\lambda^* = 1, 2$ with discounted reward criterion. Details of these methods are given in Appendix B. They have also been used in Example 2.

3.5 Equilibrium Sets: $M/M/1$ queues

In this section, we show that for a $M/M/1$ queue with same costs and parameters and among QoS measure (L and L_1) and cost criteria (discounted and average), changing the cost criteria while fixing the QoS measure (and vice versa), leads to a change in equilibrium behaviour. Also, we show that the equilibrium set in $M/M/1$ queue are due to multiple optima.

Example 1 ($M/M/1$ **queue, QoS measure** L**, Discounted vs Average cost criteria).** *For* $M/M/1$ *queue with QoS measure* L*, using average cost criteria can give rise to equilibrium set whereas discounted cost criteria can give rise to equilibrium point. This difference can be seen in Figure 4 for system costs and parameters as* $h = 25$*,* $r = 10$*,* $\mu = 10$*,* $\alpha = 0.5$*,* $m = 4.7$ *and* $e = 5$*.*

Using Theorem 7, it was observed that at $\lambda = 4.1578$ *there exist multiple optimal control limits* 3 *and* 2 *and* $l_0 = 4.1578$ *defined in Definition 1 (see Figure 3 also).*

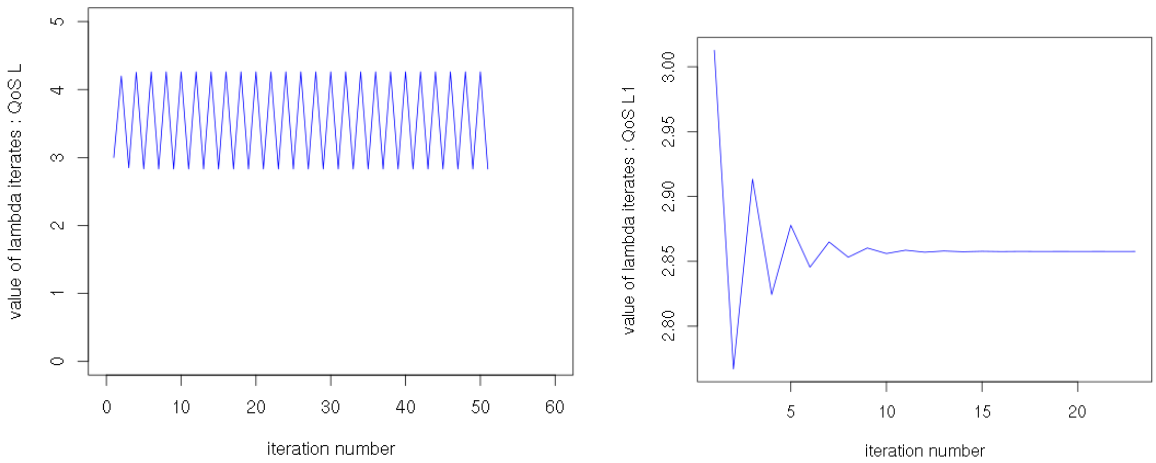


(a) (b)

Figure 4: For a $M/M/1$ queue with costs and parameters as given in Example 1 and QoS measure L : (a) Use of average cost criteria leads to equilibrium set $E = E_1 \cup E_2$ where $E_1 = (0, 4.157461]$ and $E_2 = [4.49077, 4.7)$ and repelling set $R = (4.157461, 4.49077)$ when $l_0 = 4.1578$, (b) Use of discounted cost criteria leads to equilibrium point at $\lambda = 4.446021$.

Example 2 ($M/M/1$ **queue, Discounted cost criteria, QoS measure** L **vs** $L1$**).** *For* $M/M/1$ *queue with discounted cost criteria, using QoS measure* L *can give rise to equilibrium set whereas QoS* *measure L_1 can give rise to equilibrium point. This difference in equilibrium behaviour can be seen in Figure 5 for system costs and parameters as $h = 1$, $r = 0.4$, $\mu = 5.5$, $\alpha = 0.54$, $m = 5$ and* $e = 4.97$.

In this example, the equilibrium set is due to multiple optimal control limits defined in Section 3.4. The control limits $R_\lambda = 1, 2$ are both optimal at $\lambda = 3.89$ which is l_0 as in Definition 1 (see Figure 3 also). That this particular λ is indeed a point of multiple optima has been verified in Appendix Section B.4 where the methods to compute the point of multiple optima are given.



(a) (b)

Figure 5: For a $M/M/1$ queue with costs and parameters as given in Example 2 and discounted cost criteria: (a) QoS measure L leads to equilibrium set $E = E_1 \cup E_2$ where $E_1 = (0, 3.7]$ and $E_2 = [4.4, 5)$ and repelling set is $R = (3.7, 4.4)$, (b) QoS measure L_1 leads to equilibrium point at $\lambda = 2.857543$.

We now have the following result whose proof is deferred to Appendix C.

Theorem 8. *For an admission controlled $M/M/1$ queue, the equilibrium sets arise due to multiple optimal control limits irrespective of the cost criteria (discounted or average) and QoS measure (L or L_1).*

Remark 6. *As mentioned above, such multiple control limits are consecutive integers; however, equilibrium sets in some $GI/M/1$ queues that we consider below are such that the control limits to the left and right of discontinuity point need not be consecutive integers (see Example 8).*

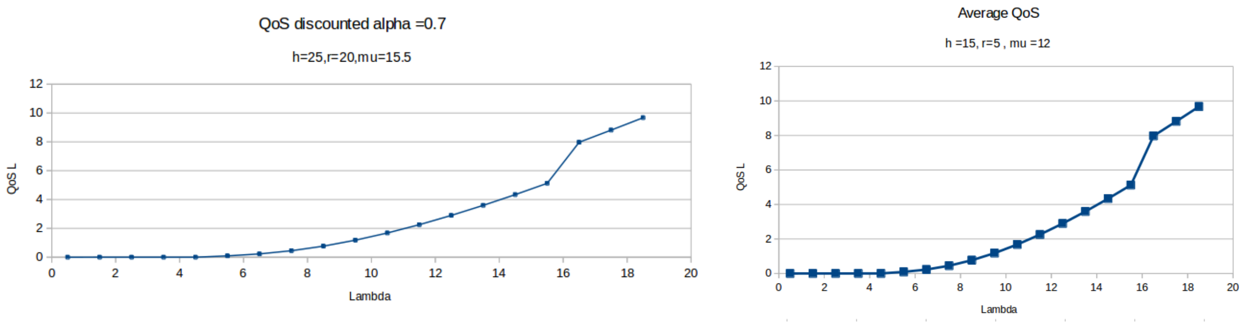
In the subsequent sections, we consider the arrival distribution to be deterministic and also generalize it to a specific discrete support distribution.

3.6 Equilibrium sets: $D/M/1$ queue

In this section we observe a queue where, customers follow deterministic arrival rate i.e., one arrival every $\frac{1}{\lambda}$ time units. Here, each λ real is the support (i.e., the support is singleton). In case of discounted cost criteria, Van Nunen and Puterman algorithm is used for calculating the control limit with expected holding cost $c(s)$ as given below :

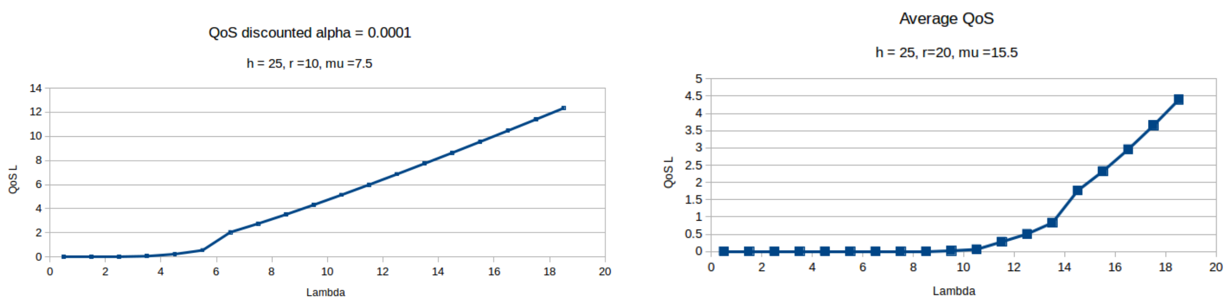
$$c(s) = \int_0^{\frac{1}{\lambda}} e^{-\alpha t} \sum_{k=0}^{s-1} h(s-k) \frac{e^{-\mu t}(\mu t)^k}{k!} dt$$

In case of average cost criteria, we use the same method as given in Section 3.2 after making appropriate changes in $c(s)$. Figure 6, 7, 8, 9 illustrates the monotonic nature of QoS in case of discounted and average cost criteria.



(a) (b)

Figure 6: QoS measure L, is monotonically increasing for a $D/M/1$ queue with costs and parameters as : (a) $h = 25$, $r = 20$, $\mu = 15.5$, $\alpha = 0.7$ in discounted settings, (b) $h = 15$, $r = 5$, $\mu = 12$ in average settings.

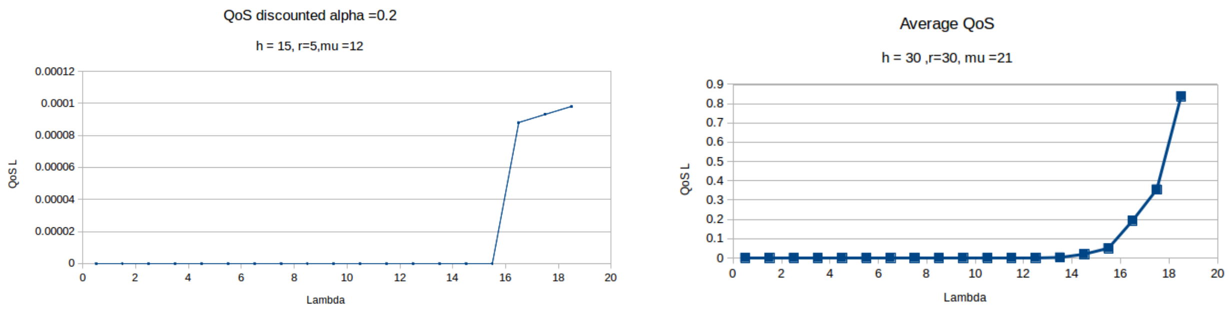


(a) (b)

Figure 7: QoS measure L, is monotonically increasing for a $D/M/1$ queue with costs and parameters as : (a) $h = 25$, $r = 10$, $\mu = 7.5$, $\alpha = 0.0001$ in discounted settings, (b) $h = 25$, $r = 20$, $\mu = 15.5$ in average settings.

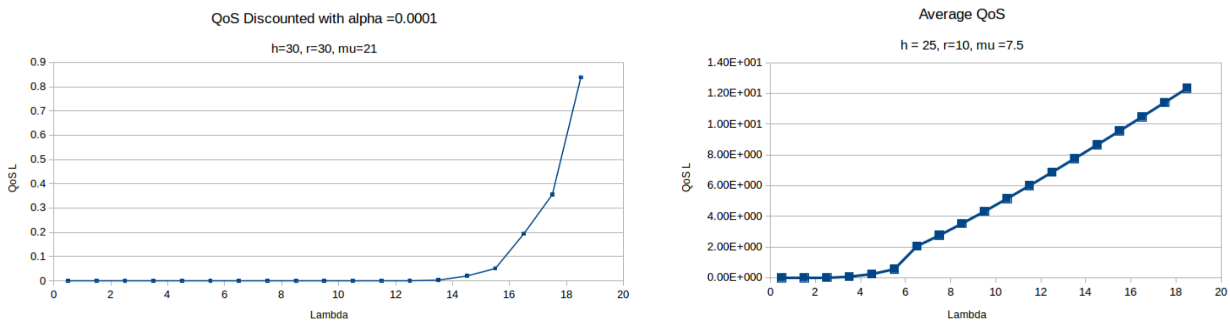
Following examples give an insight in the change in equilibrium behaviour due to change in cost criteria and change in QoS measure.

Example 3 ($D/M/1$ **queue, QoS measure** L_1**, Discounted vs Average cost criteria).** *For* $D/M/1$ *queue with QoS measure* $L1$*, using average cost criteria can give rise to equilibrium point whereas*



(a) (b)

Figure 8: QoS measure L, is monotonically increasing for a $D/M/1$ queue with costs and parameters as : (a) $h = 15$, $r = 5$, $\mu = 12$, $\alpha = 0.2$ in discounted settings, (b) $h = 30$, $r = 30$, $\mu = 21$ in average settings.

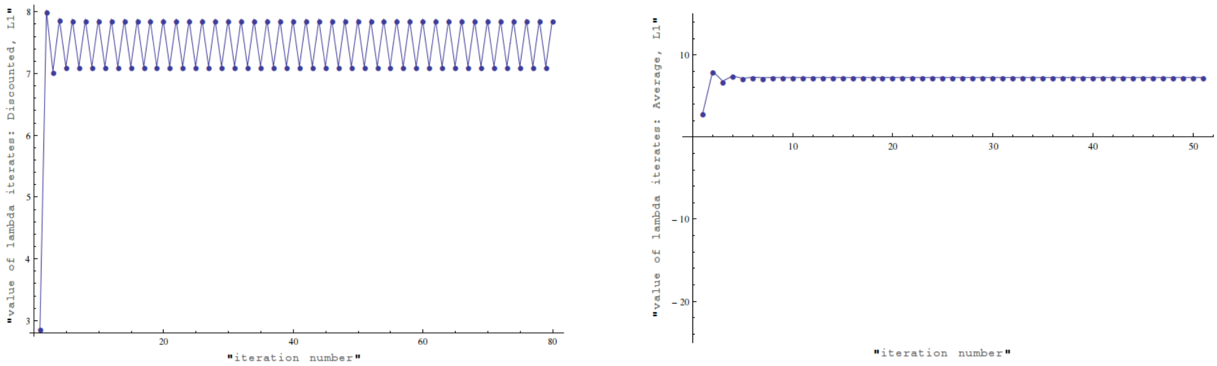


(a) (b)

Figure 9: QoS measure L, is monotonically increasing for a $D/M/1$ queue with costs and parameters as : (a) $h = 30$, $r = 30$, $\mu = 21$, $\alpha = 0.0001$ in discounted settings, (b) $h = 25$, $r = 10$, $\mu = 7.5$ in average settings.

discounted cost criteria can give rise to equilibrium set. This difference can be seen in Figure 10 for system costs and parameters $h = 15$, $r = 5$, $\mu = 12$, $\alpha = 0.68$, $m = 8$ and $e = 2$.

Using Theorem 7, we conclude that at $\lambda = 7.49299$ there exists multiple optimal control limits 3 and 2 which leads to equilibrium sets and $l_0 = 7.49299$ as defined in Definition 1 (Figure 3).



(a) (b)

Figure 10: For a $D/M/1$ queue with costs and parameters as given in Example 3 and QoS measure $L1$: (a) Use of discounted cost criteria leads to equilibrium set $E = E_1 \cup E_2$ where $E_1 = (0, 7.204]$ and $E_2 = [7.84, 8)$ and repelling set $R = (7.20, 7.84)$, (b) Use of average cost criteria leads to equilibrium point at $\lambda = 7.345$.

Example 4 ($D/M/1$ **queue, Discounted cost criteria, QoS measure** L **vs** $L1$)**.** *For $D/M/1$ queue with discounted cost criteria, using QoS measure L_1 can give rise to equilibrium set whereas using QoS measure L can give rise to equilibrium point. This difference is depicted in Figure 11 for system costs and parameters as $h = 5$, $r = 3$, $\mu = 7.5$, $\alpha = 0.8$, $m = 8.8$ and $e = 1$.*

Using Theorem 7, we conclude that at $\lambda = 7.70146$ there exists multiple optimal control limits 3 and 2 which leads to equilibrium set with $l_0 = 7.70146$.

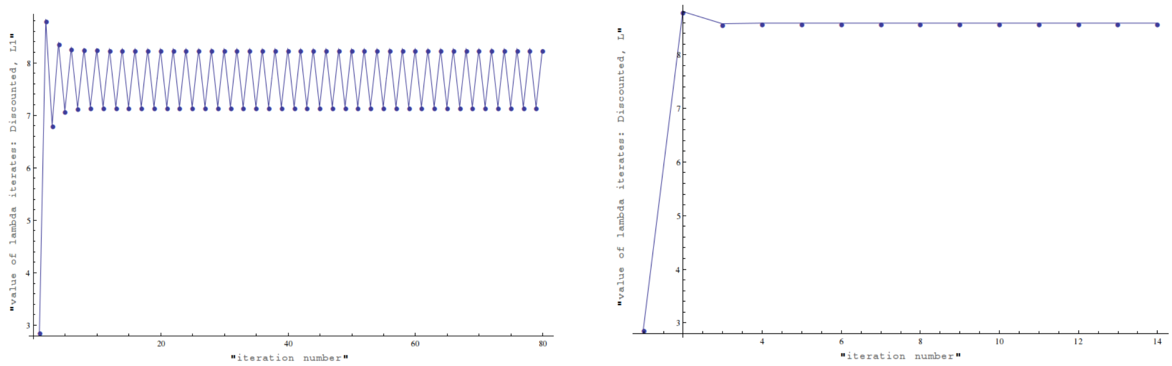
Using the same arguments as in the proof of Theorem 8, we have the following result:

Theorem 9. *For an admission controlled $D/M/1$ queue, the equilibrium sets can arise due to multiple optimal control limits irrespective of the cost criteria (discounted or average) and QoS measure (L or L_1).*

3.7 Equilibrium sets: Discrete arrival distribution

In this section, we consider the case when number of customers arrived per unit time follow a discrete distribution of the form given below:

$$G(\cdot) = \begin{cases} u_1 & \text{when arrival rate is } \lambda_1 \\ u_2 & \text{when arrival rate is } \lambda_2 \\ u_3 & \text{when arrival rate is } \lambda_3 \end{cases}$$



(a) (b)

Figure 11: For a $D/M/1$ queue with costs and parameters as given in Example 4 and discounted cost criteria: (a) QoS measure L_1 leads to equilibrium set $E = E_1 \cup E_2$ where $E_1 = (0, 7.35]$ and $E_2 = [8.05, 8.8)$ and repelling set $R = (7.35, 8.05)$, (b) QoS measure L leads to equilibrium point at $\lambda = 8.554$.

where $G(\cdot)$ is the probability mass function of random variable denoting number of arrivals per unit time. Using a specific choice of probabilities $\{u_i\}_1^3$ and support set $\{\lambda_i\}_1^3$ we make sure that the arrival rate λ is locally continuous over $(0, m)$. The details to generate such arrival rates, along with an example, are given in Appendix D. In case of discounted cost criteria, Van Nunen and Puterman algorithm is used for calculating the control limit with $c(s)$ as given below :

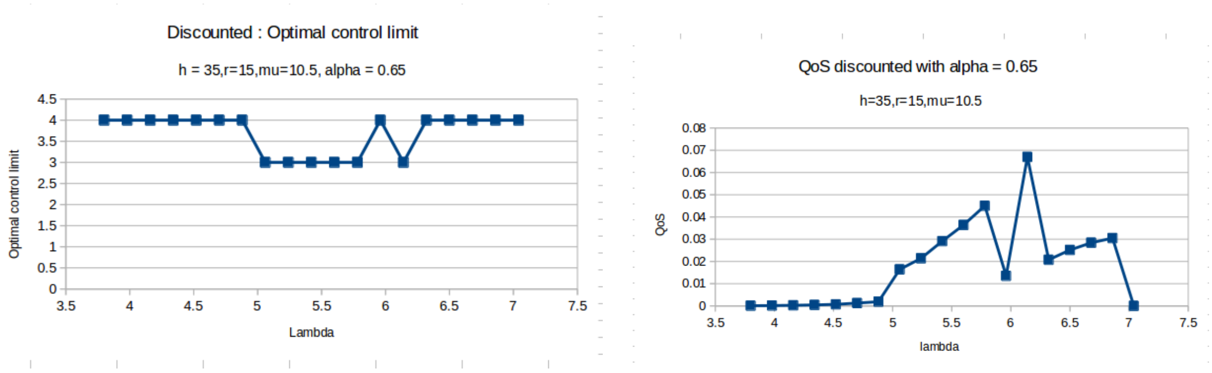
$$c(s) = \sum_{i=1}^{3} u_i \int_0^{\frac{1}{\lambda_i}} e^{-\alpha t} \sum_{k=0}^{s-1} h(s-k) \frac{e^{-\mu t}(\mu t)^k}{k!} dt$$

To compute the optimal control limit for average cost criteria, we use the method as given in Section 3.2 after making appropriate changes in $c(s)$.

Optimal control limit was seen to be non-monotonic with respect to λ as can be seen in Table 1. Due to this, QoS measures L and L_1 are also non-monotonic with respect to λ (Example 5 and 6). Hence, we cannot assume that $A1$ holds; however, equilibrium sets given in Definition 1 can exist as in Example 7 and 8. Also, the equilibrium behaviour for these queues can be quite complicated; see Section 3.7.2.

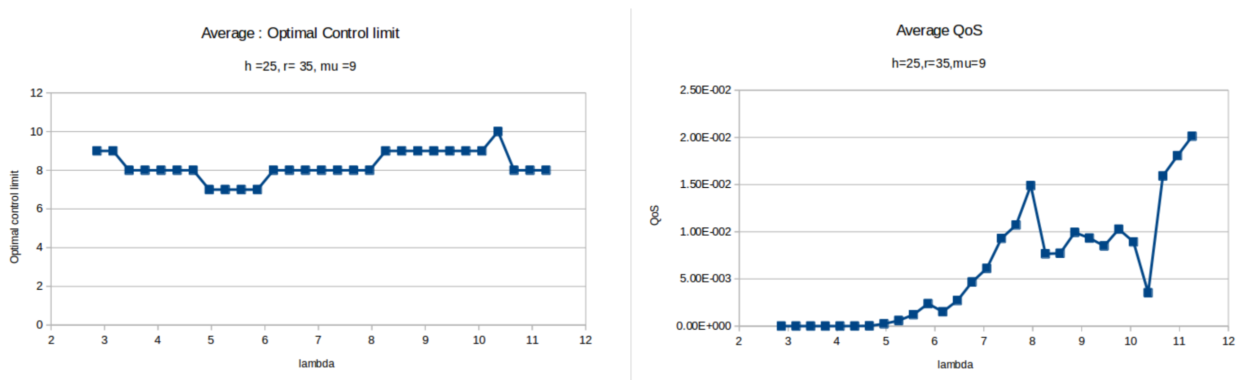
Example 5 **(Discrete support, QoS measure L, Discounted cost criteria, Non-monotonic nature).** *For a queue with discrete arrival rate distribution, using discounted cost criteria, the optimal control limit and the QoS measure L are not monotonic w.r.t. λ, as shown in Figure 12 for system cost and parameters as $h = 35$, $r = 15$, $\mu = 10.5$, $\alpha = 0.65$.*

Example 6 **(Discrete support, QoS measure L, Average cost criteria, Non-monotonic nature).** *For a queue with discrete arrival rate distribution, using average cost criteria, the optimal control limit and the QoS measure L are not monotonic w.r.t. λ, as shown in Figure 13 for system cost and parameters as $h = 25$, $r = 35$, $\mu = 9$.*



(a) (b)

Figure 12: For a discrete arrival rate distribution queue with costs and parameters as given in Example 5 and discounted cost criteria (a) shows non-monotonic nature of optimal control limit and (b) shows non-monotonic nature QoS measure L.



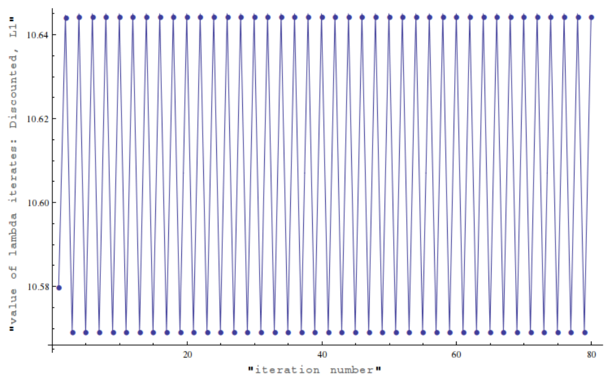
(a) (b)

Figure 13: For a discrete arrival rate distribution queue with costs and parameters as given in Example 6 and average cost criteria (a) shows non-monotonic nature of optimal control limit and (b) shows non-monotonic nature QoS measure L.

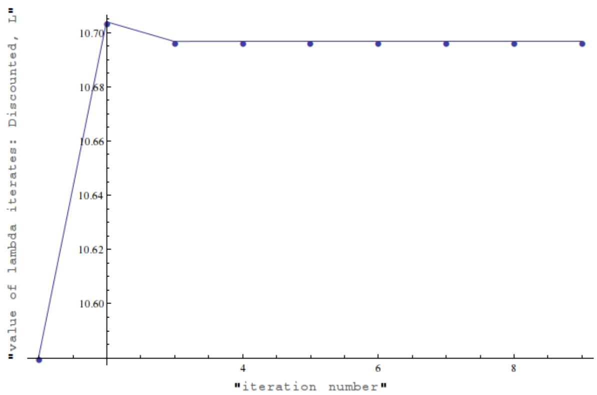
3.7.1 Numerical examples for queues with discrete arrival rate distribution

Example 7 **(Discrete support, Discounted cost criteria, QoS measure L vs L_1).** *For a queue with discrete arrival rate distribution with discounted cost criteria, QoS measure L_1 leads equilibrium set $E = E_1 \cup E_2$ whereas using QoS measure L makes the λ iterates to converge to a unique equilibrium point. This difference can be seen in Figure 14 for system costs and parameters as $h = 5$, $r = 3$, $\mu = 10$, $\alpha = 0.9$, $m = 10.71$ and $e = 5$.*

Also, from Figure 15 we identify the equilibrium set $E = E_1 \cup E_2$ where $E_1 = (2.85, 10.569]$ and $E_2 = [10.644, 10.71)$ and repelling set $R = (10.569, 10.644)$. We observe that this equilibrium set is due to multiple optimal control limit. However, there is a change of support from $\{1, 16, 17\}$ to $\{2, 17, 18\}$ at this point as seen in Table 1. At $l_0 \in R$, there is downward discontinuity in $f(\lambda)$ due to multiple optimal control limits 7 and 6. Also, we have observed that change in support leads to upward discontinuity in $f(\lambda)$. In this example, at l_0, overall effect of multiple optimal control limits and change in support has a downward discontinuity; leading to existence of equilibrium set consistent with Definition 1.



(a)

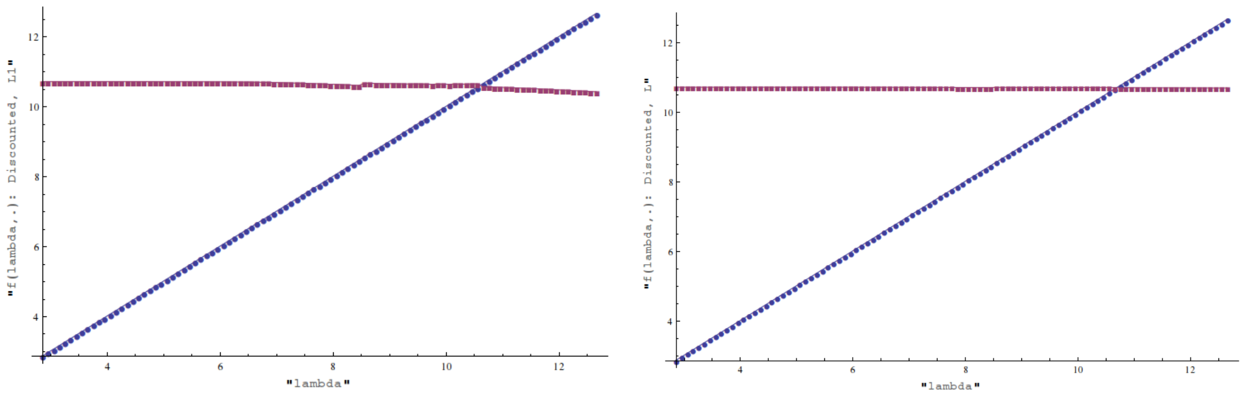


(b)

Figure 14: For a queue with costs and parameters as given in Example 7 and discrete arrival distribution with discounted cost criteria: (a) QoS measure L_1 leads to an equilibrium set $E = E_1 \cup E_2$ where $E_1 = (2.85, 10.569]$ and $E_2 = [10.644, 10.71)$ and repelling set $R = (10.569, 10.644)$, (b) QoS measure L leads to an equilibrium point 10.6965.

Example 8 **(Discrete support, Average cost criteria, QoS measure L vs L_1).** *For a queue with discrete arrival rate distribution with average cost criteria, QoS measure L_1 can give rise to an equilibrium set $E = E_1 \cup E_2$ whereas QoS measure L can give rise to an equilibrium point. The above difference can be seen in Figure 16 for system costs and parameters as $h = 25$, $r = 35$, $\mu = 9$, $m = 10.65$ and $e = 9$.*

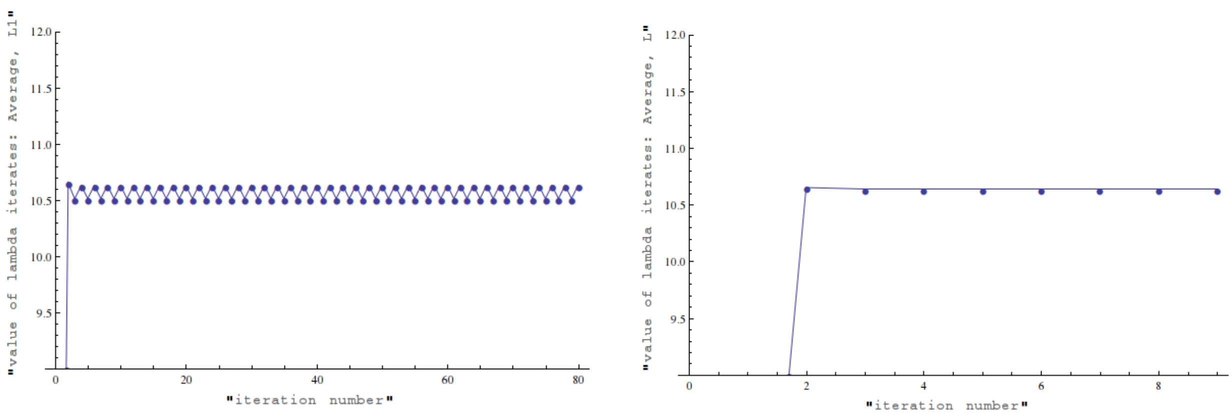
Also, from Figure 17 we identify the equilibrium set $E = E_1 \cup E_2$ where $E_1 = (2.85, 10.5089]$, $E_2 = [10.6254, 10.65)$, and repelling set $R = (10.5089, 10.6254)$. There is also a change of support from $\{1, 16, 17\}$ to $\{2, 17, 18\}$ at $\lambda = 10.6$ as seen in Table 1.



(a) (b)

Figure 15: For a queue with costs and parameters as given in Example 7 and discounted cost criteria : (a) $f(\lambda, L_1)$ vs λ shows that the equilibrium set is $E = E_1 \cup E_2$ where $E_1 = (2.85, 10.569]$ and $E_2 = [10.644, 10.71)$ and repelling set is $R = (10.569, 10.644)$ (b) $f(\lambda, L)$ vs λ shows that the equilibrium point is 10.6965.

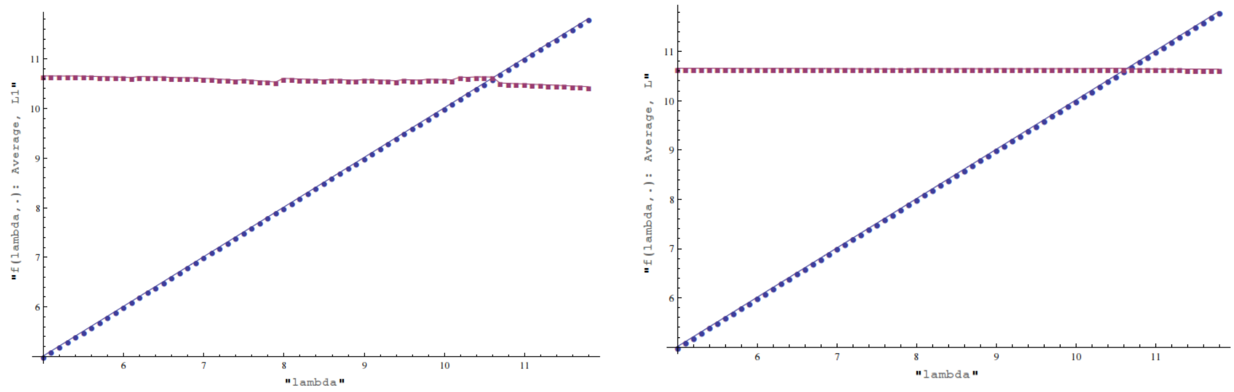
Another observation in this example is that the control limit for two consecutive intervals is not consecutive (for $\lambda \in (10.375, 10.6],\ R^*_\lambda = 10$ and for $\lambda \in (10.6, 11.136],\ R^*_\lambda = 8$); for an average cost criteria $M/M/1$ queue Haviv and Puterman (7) show that there can be at most two optimal control limits. This behaviour of optimal policies (control limit) can be attributed to change in support.



(a) (b)

Figure 16: For a queue with costs and parameters as given in Example 8 and discrete arrival distribution with average cost criteria: (a) QoS measure L_1 leads to equilibrium set $E = E_1 \cup E_2$ where $E_1 = (2.85, 10.5089]$, $E_2 = [10.6254, 10.65)$, and $R = (10.5089, 10.6254)$ (b) QoS measure L leads to a unique equilibrium point 10.63667.

Example 9 (**Discrete support, Discounted cost criteria, QoS measure** L **vs** L_1). *For a queue with discrete arrival rate distribution with discounted cost criteria, using QoS measure L_1 leads to some notion of equilibrium set depending upon the initial λ whereas using QoS measure L makes the λ*



(a) (b)

Figure 17: For a queue with costs and parameters as given in Example 8 and average cost criteria: (a) $f(\lambda, L_1)$ vs λ shows that the equilibrium set $E = E_1 \cup E_2$ where $E_1 = (2.85, 10.5089]$, $E_2 = [10.6254, 10.65)$, and repelling set is $R = (10.5089, 10.6254)$, (b) $f(\lambda, L)$ against λ shows that the equilibrium point is 10.63677.

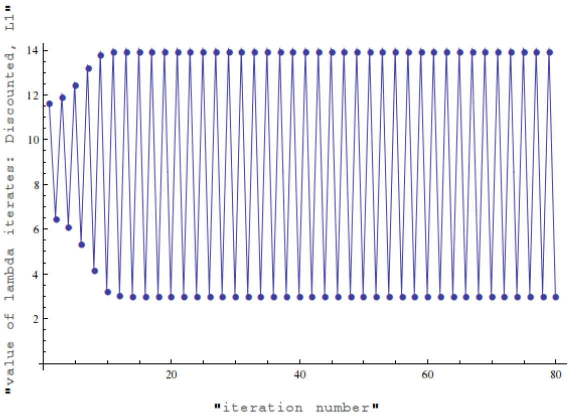
iterates to converge to a unique equilibrium point. This difference can be seen in Figure 18 for system costs and parameters as $h = 25$, $r = 10$, $\mu = 7.5$, $\alpha = 1.3$, $m = 14$ and $e = 9$.

Even though $A1$ does not hold, we can still identify equilibrium set $E = E_1 \cup E_2$ where $E_1 = (2.85, 6.775]$ and $E_2 = [11.69, 14)$ which are consistent with Definition 1. Instead of repelling set R in between them as in $M/M/1$ and $D/M/1$ queues, there exists an invariant/equilibrium set $E_0 = (6.775, 11.69)$ around an equilibrium point 8.7157. Also, it was observed that initial value of λ plays an important role in deciding which regime of λ the system is going to remain in. We also plot $f(\cdot)$ (with QoS measure L_1 and L), against λ in Figure 19 for this example. It can be clearly seen from Figure 19 (a), that the non-monotonic nature of $f(\cdot)$ caused due to change in support of arrival rate leads to equilibrium sets E and E_0.

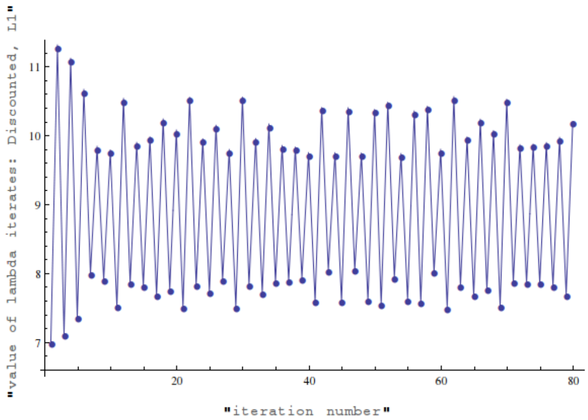
Example 10 (Discrete support, Average cost criteria, QoS measure L vs L_1). *For a queue with discrete arrival rate distribution with average cost criteria, using QoS measure L_1 leads to some notion of equilibrium set depending upon the initial λ whereas QoS measure L can give rise to equilibrium point. The above difference can be seen in Figure 20 for system costs and parameters as $h = 35$, $r = 15$, $\mu = 10.6$, $m = 14$ and $e = 10$.*

From (a) of Figure 21, we see that assumption $A1$ does not hold. We evaluated $f(\lambda, L_1)$ over $(2.86, 16.5)$ in increments of 0.2 and it was noted that for $\lambda \in (2.85, 7.21]$, $f(\lambda) \in [11.725, 14]$ and vice versa. This has been shown in Table 2. So, we can identify equilibrium set $E = E_1 \cup E_2$ where $E_1 = (0, 7.21]$ and $E_2 = [11.725, 14]$. Also, there exist an invariant/equilibrium set $E_0 = (7.21, 11.275)$ around an equilibrium point 9.158178. From Table 1 we can see that for this $GI/M/1$ queue, the control limit is constant; so, the equilibrium sets are due to change in the support.

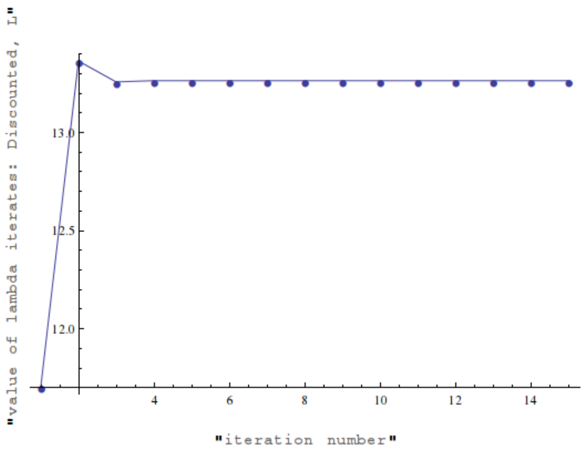
Example 11 (Discrete support, Discounted cost criteria, QoS measure L vs L_1). *For a queue with discrete arrival rate distribution with discounted cost criteria, using QoS measure L makes the λ*



(a)

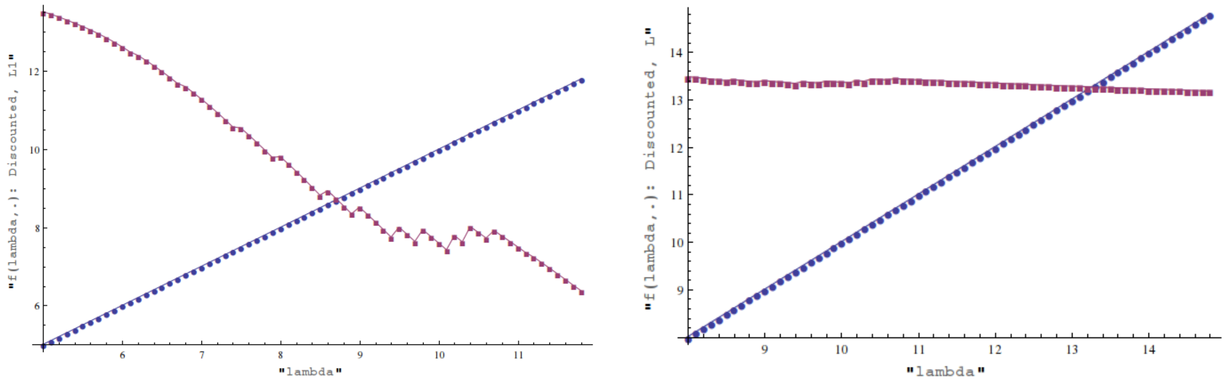


(b)



(c)

Figure 18: For a queue with costs and parameters as given in Example 9 and discrete rate arrival distribution with discounted cost criteria: (a) QoS measure L_1 leads to toggling of iterates between $E_1 = (2.85, 6.775]$ and $E_2 = [11.69, 14)$ when the initial $\lambda_0 \in E_1$ or $\lambda_0 \in E_2$ (b) QoS measure L_1 leads to non convergent iterates of λ if initial $\lambda_0 \in E_0 := (6.775, 11.69)$; E_0 is an equilibrium/invariant set (c) QoS measure L leads to a unique equilibrium point 13.2606.



(a) (b)

Figure 19: For a queue with costs and parameters as given in Example 9 and discounted cost criteria : (a) $f(\lambda, L_1)$ vs λ shows that the equilibrium sets are $E = E_1 \cup E_2$ where $E_1 = (2.85, 6.775]$ and $E_2 = [11.69, 14)$ and $E_0 = (6.775, 11.69)$(b) $f(\lambda, L)$ vs λ shows that the equilibrium point is 13.2606.

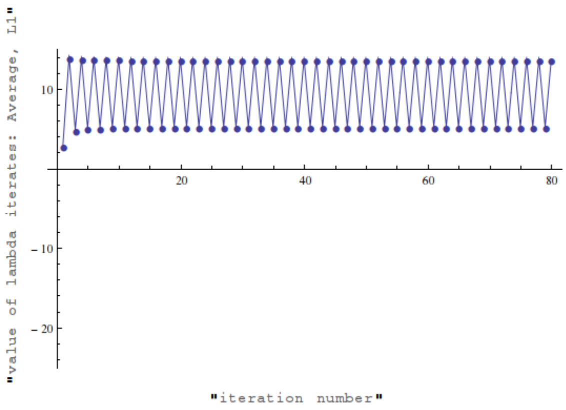
iterates to converge to a unique equilibrium point whereas using QoS measure L_1 leads to multiple (4) equilibrium points depending upon the initial λ. This difference can be seen in Figure 22 for system costs and parameters as $h = 35$, $r = 15$, $\mu = 10.5$, $\alpha = 0.68$, $m = 14$ and $e = 18$.

It can be observed from Figure 23 that assumption $A1$ does not hold for this example. However as seen in Table 1, increase in control limit from 3 to 4 (multiple optimal control limits) when support changes from $\{1, 10, 11\}$ to $\{1, 11, 12\}$ induces a non monotonicity in f along with discontinuity. There are non monotonicities in f later also due to change in support. This leads to more than one equilibrium point and no equilibrium set. However, as can be seen in Figure 22, there exist set $\tilde{E} = \tilde{E}_1 \cup \tilde{E}_2 \cup \tilde{E}_3$ where $\tilde{E}_1 = [10.40991425, 10.42658945]$, $\tilde{E}_2 = [10.58613254, 10.62989617]$ and $\tilde{E}_3 = [10.87676771, 10.90933302]$ such that the sequence $\{\lambda_i\}_{i\geq 0}$ lies in $\tilde{E}_1$, $\tilde{E}_2$ and $\tilde{E}_3$ infinitely often respectively. These have been verified by starting from different initial values of λ. We also have a set $\tilde{R} = (0, m) \setminus \tilde{E}$ such that if $\lambda \in \tilde{R}$ then $f^n(\lambda) \in \tilde{E}$ for some finite n. A general framework for this behaviour has been given in Section 3.7.2.

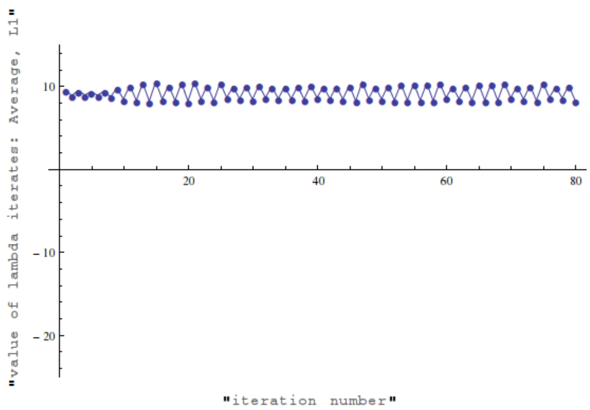
Example 12 (Discrete support, Discounted cost criteria, QoS measure L vs L_1). *For a queue with discrete arrival rate distribution with discounted cost criteria, using QoS measure L makes the λ iterates to converge to a unique equilibrium point whereas using QoS measure L_1 leads to neither an equilibrium point nor an equilibrium set. This difference can be seen in Figure 24 for system costs and parameters as $h = 5$, $r = 3$, $\mu = 6$, $\alpha = 0.9$, $m = 14$ and $e = 5$.*

Clearly, from Figure 25 there does not exist an equilibrium point. Even though the discontinuity in the f which decides the equilibrium behaviour is due to multiple optimal control limit, there does not exist the equilibrium set as defined in Section 2.1.

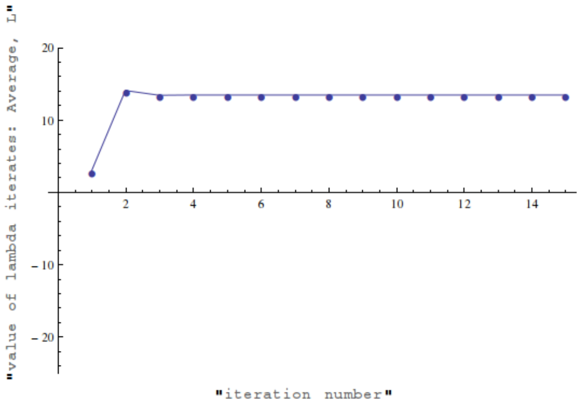
However, there exist set $\tilde{E} = \tilde{E}_1 \cup \tilde{E}_2 \cup \tilde{E}_3$ where $\tilde{E}_1 = [7.713797269, 8.037336962]$, $\tilde{E}_2 = [9.828436485, 10.1945525]$ and $\tilde{E}_3 = [11.56635919, 11.89060517]$ such that the sequence $\{\lambda\}_{i\geq 0}$



(a)

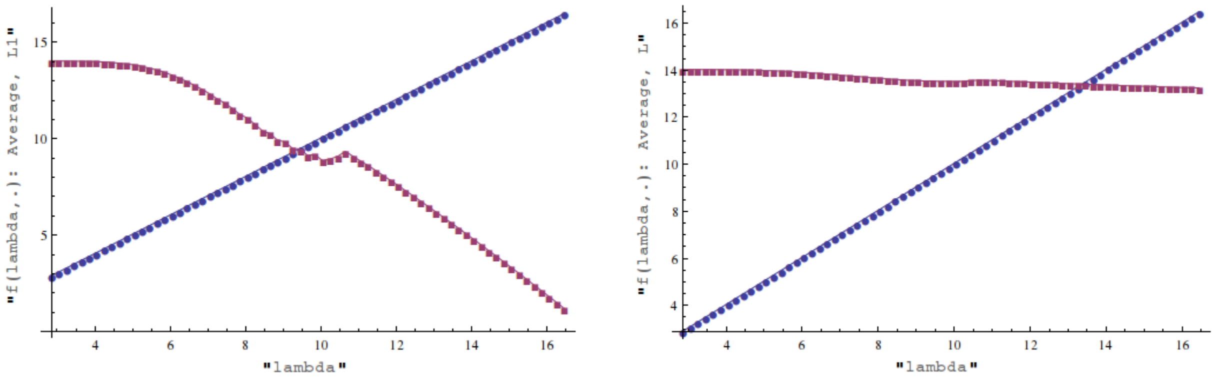


(b)



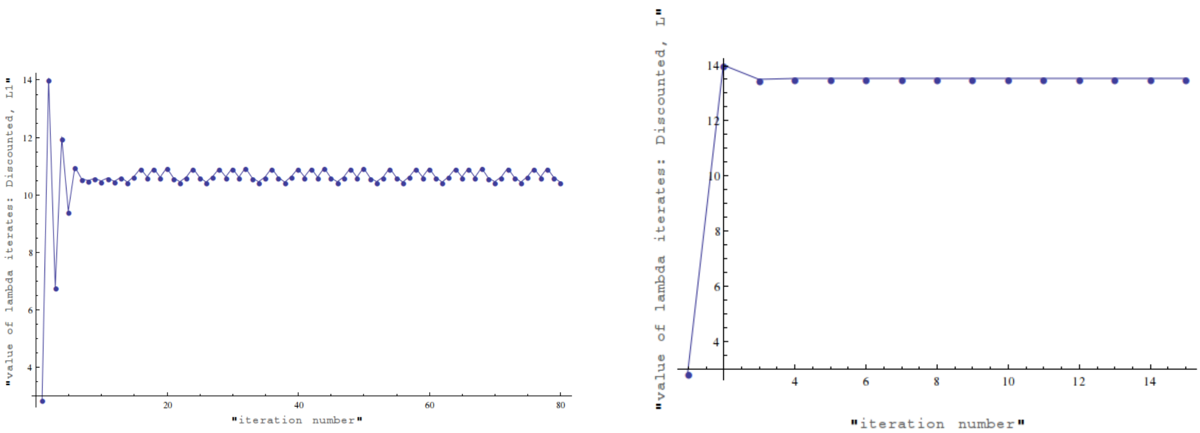
(c)

Figure 20: For a queue with costs and parameters as given in Example 10 and discrete arrival distribution with average cost criteria:(a) QoS measure L_1 leads to toggling of iterates between $E_1 = (2.85, 7.21]$ and $E_2 = [11.725, 14)$ when the initial $\lambda_0 \in E_1$ or $\lambda_0 \in E_2$ (b) QoS measure L_1 leads to non convergent iterates of λ if initial $\lambda_0 \in E_0 := (7.21, 11.725)$; E_0 is an equilibrium/invariant set, (c) QoS measure L leads to a unique equilibrium point 13.317.



(a) (b)

Figure 21: For a queue with costs and parameters as given in Example 10 and average cost criteria: (a) $f(\lambda, L_1)$ vs λ shows that the equilibrium sets are $E = E_1 \cup E_2$ where $E_1 = (2.85, 7.21]$ and $E_2 = [11.725, 14)$ and $E_0 = (7.21, 11.725)$, (b) $f(\lambda, L)$ against λ shows that the equilibrium point is 13.317.



(a) (b)

Figure 22: For a queue with costs and parameters as given in Example 11 and discrete arrival rate distribution with discounted cost criteria: (a) QoS measure L_1 leads to 4 equilibrium points $8.3998, 10.3489, 10.5223$ and 10.7666 along with set $\tilde{E}$ and $\tilde{R}$ as given the description of example. (b) QoS measure L leads to a unique equilibrium point 13.4521.

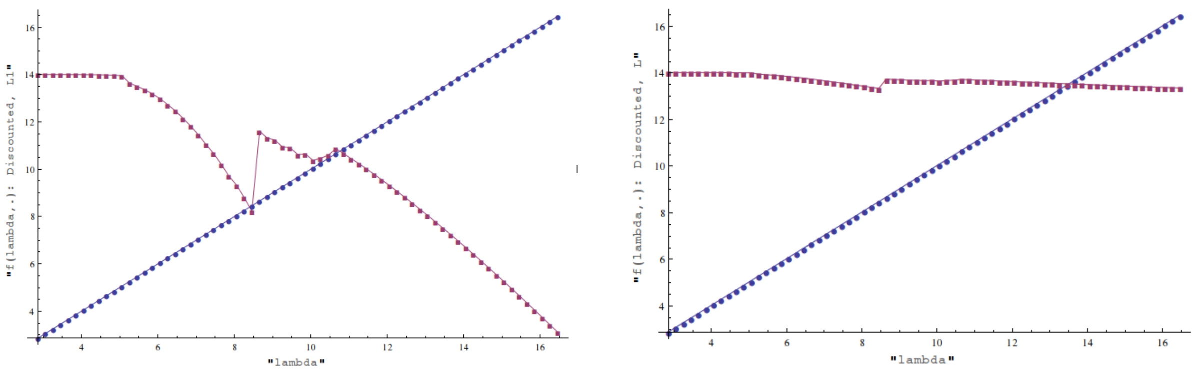
S. no	Discrete support	Arrival rate intervals	$R^*_{\lambda,dc}$ (Ex 7)	$R^*_{\lambda,av}$ (Ex 8)	$R^*_{\lambda,dc}$ (Ex 9)	$R^*_{\lambda,av}$ (Ex 10)	$R^*_{\lambda,dc}$ (Ex 11)	$R^*_{\lambda,dc}$ (Ex 12)
1	1, 3, 4	(2.85, 3.775]	7, 6	9, 8	3	3	4	4, 3
2	1, 4, 5	(3.775, 4.6]	6	8, 7	3	3	4	3
3	1, 5, 6	(4.6, 5.375]	6	8, 7	3	3	4, 3	3
4	1, 6, 7	(5.375, 6.1]	6	7	3	3	3	3
5	1, 7, 8	(6.1, 6.775]	6	8	3	3	3	3
6	1, 8, 9	*(6.775, 7.4]*	6	8	*3*	*3*	3	3
7	1,9, 10	*(7.4, 7.975]*	6	8	*3*	*3*	3	3
8	1, 10, 11	*(7.975, 8.5]*	6	9	*3*	*3*	*3*	4
9	1, 11, 12	*(8.5, 8.975]*	7	9	*3*	*3*	*4*	4
10	1, 12, 13	*(8.975, 9.4]*	6	9	*3*	*3*	*4*	4
11	1, 13, 14	*(9.4, 9.775]*	6	9	*3*	*3*	*4*	4
12	1, 14, 15	*(9.775, 10.1]*	6	9	*3*	*3*	*4*	4
13	1, 15,16	*(10.1, 10.375]*	6	10	*3*	*3*	*4*	*4*
14	1, 16, 17	*(10.375, 10.6]*	*7*	*10*	*3*	*3*	*4*	*3*
15	2, 17, 18	*(10.6, 11.136]*	*6*	*8*	*3*	*3*	*4*	3
16	2, 18, 19	*(11.136, 11.69]*	6	8	*3*	*3*	4	3
17	2, 19, 20	*(11.69, 12.26]*	6	8	3	*3*	4	3
18	2, 20, 21	(12.26, 12.83]	6	8	3	3	4	3
19	2, 21, 22	(12.83, 13.4]	6	8	3	3	4	3
20	2, 22, 23	(13.4, 13.97]	6	8	3	3	4	3
21	2, 23, 24	(13.97, 14.54]	6	8	3	3	4	3
22	2, 24, 25	(14.54, 15.11]	6	8	3	3	4	3
23	2, 25, 26	(15.11, 15.68]	6	8	3	3	4	3
24	2, 26, 27	(15.68, 16.25]	6	8	3	3	4	3
25	2, 27, 28	(16.25, 16.82]	6	8	3	3	4	3, 4
26	2, 28, 29	(16.82, 17.39]	6	8	3	3	4	4

Table 1: For a queue with discrete support arrival rate distribution, the change of control limit $R^*_{\lambda,dc}$ (under discounted cost criteria) and $R^*_{\lambda,av}$ (under average cost criteria), can be seen when the support of the arrival rate changes. The text in italics represents the support and corresponding optimal control limit which affects the equilibrium behaviour.

lies in $\tilde{E}_1$, $\tilde{E}_2$ and $\tilde{E}_3$ infinitely often respectively. The boundaries of these sets have been identified by running the experiment with different initial value of $\lambda_0 = 2.86,\ 8.1,\ 12.5$. This can be seen in Table 3. We also have $\tilde{R} = (0, m) \setminus \tilde{E}$ such that if $\lambda \in \tilde{R}$ then $f^n(\lambda) \in \tilde{E}$ for some finite n. We give a general definition of these sets; generalized equilibrium sets, in the next subsection.

S. no	$\lambda_1 \in E_1$	f(λ_1) $\in E_2$	$\lambda_2 \in E_2$	f(λ_2) $\in E_1$	$\lambda_3 \in E_0$	f(λ_3) $\in E_0$
1	2.86	13.99927	11.86	7.12060	7.66	11.27339
2	3.16	13.99738	12.01	6.90093	7.81	11.02076
3	3.46	13.99310	12.16	6.67658	7.96	10.75087
4	4.21	13.95209	12.31	6.45113	8.11	10.62847
5	4.36	13.93498	12.46	6.22849	8.41	10.08194
6	4.96	13.80563	12.61	6.00152	8.56	10.00251
7	5.56	13.54276	12.76	5.77015	8.86	9.44557
8	5.71	13.45401	12.91	5.54000	9.16	9.15480
9	6.01	13.23576	13.06	5.31065	9.46	8.92531
10	6.16	13.10622	13.21	5.07723	9.91	8.51220
11	6.31	12.97141	13.36	4.83970	10.21	8.43798
12	6.46	12.82123	13.51	4.60552	10.81	8.59103
13	6.91	12.31042	13.66	4.37022	10.96	8.38316
14	7.06	12.11367	13.81	4.13111	11.11	8.16985
15	7.21	11.89977	13.96	3.88815	11.26	7.98328

Table 2: For a queue with costs and parameters as given in Example 10 and discrete arrival distribution with average cost criteria QoS measure L_1 leads to toggling of iterates between $E_1 = (2.85, 7.21]$ and $E_2 = [11.725, 14)$ when the initial $\lambda_0 \in E_1$ or $\lambda_0 \in E_2$ and non convergent iterates of λ if initial $\lambda_0 \in E_0 := (7.21, 11.725)$; E_0 is an equilibrium/invariant set,

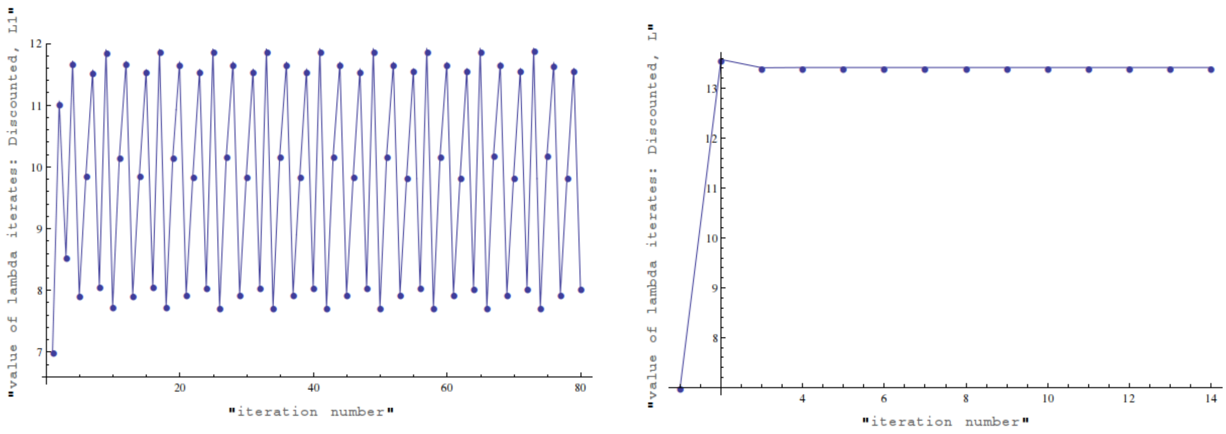


(a) (b)

Figure 23: For a queue with costs and parameters as given in Example 11 and discounted cost criteria : (a) $f(\lambda, L_1)$ vs λ shows that the there is no equilibrium set but 4 equilibrium points and generalized equilibrium sets as defined in Section 3.7.2 (b) $f(\lambda, L)$ vs λ shows that the unique equilibrium point is 13.4521.

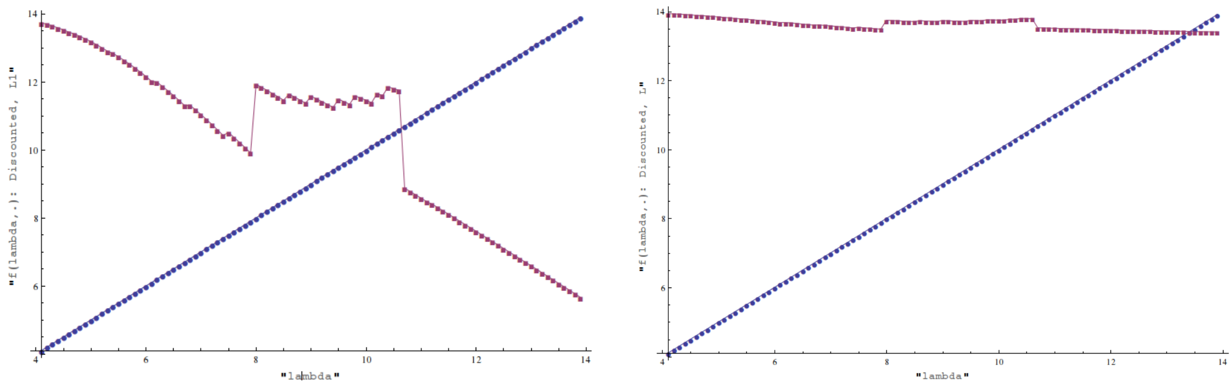
iterate no.	λ	f(λ)	λ	f(λ)	λ	f(λ)
1	2.86000	13.99784	8.10000	11.83537	12.50000	7.09862
2	13.99784	5.55645	11.83537	7.76871	7.09862	10.89012
3	5.55645	12.67555	7.76871	10.11148	10.89012	8.68614
4	12.67555	6.91956	10.11148	11.70007	8.68614	11.56079
5	6.91956	11.14559	11.70007	7.90198	11.56079	8.04296
6	11.14559	8.45244	7.90198	9.90459	8.04296	11.88569
7	8.45244	11.50473	9.90459	11.51706	11.88569	7.71869
8	11.50473	8.09930	11.51706	8.08693	7.71869	10.18720
9	8.09930	11.83599	8.08693	11.84697	10.18720	11.65693
10	11.83599	7.76809	11.84697	7.75720	11.65693	7.94555
11	7.76809	10.11243	7.75720	10.12900	7.94555	9.83531
12	10.11243	11.69953	10.12900	11.69013	9.83531	11.56194
13	11.69953	7.90250	11.69013	7.91169	11.56194	8.04179
14	7.90250	9.90376	7.91169	9.88922	8.04179	11.88671
15	9.90376	11.51760	9.88922	11.52707	11.88671	7.71768
16	11.51760	8.08639	11.52707	8.07689	7.71768	10.18872
17	8.08639	11.84745	8.07689	11.85586	10.18872	11.65606
18	11.84745	7.75672	11.85586	7.74837	11.65606	7.94644
19	7.75672	10.12972	7.74837	10.14240	7.94644	9.83389
20	10.12972	11.68972	10.14240	11.68252	9.83389	11.56285
21	11.68972	7.91210	11.68252	7.91946	11.56285	8.04087
22	7.91210	9.88856	7.91946	9.87689	8.04087	11.88752
23	9.88856	11.52750	9.87689	11.53508	11.88752	7.71687
24	11.52750	8.07646	11.53508	8.06884	7.71687	10.18993
25	8.07646	11.85624	8.06884	11.86297	10.18993	11.65536
26	11.85624	7.74799	11.86297	7.74131	11.65536	7.94715
27	7.74799	10.14297	7.74131	10.15309	7.94715	9.83276
28	10.14297	11.68219	10.15309	11.67643	9.83276	11.56358
29	11.68219	7.91979	11.67643	7.92567	11.56358	8.04014
30	7.91979	9.87636	7.92567	9.86701	8.04014	11.88816
31	9.87636	11.53542	9.86701	11.54147	11.88816	7.71624
32	11.53542	8.06850	11.54147	8.06240	7.71624	10.19089
33	8.06850	11.86327	8.06240	11.86863	10.19089	11.65481
34	11.86327	7.74101	11.86863	7.73567	11.65481	7.94771
35	7.74101	10.15354	7.73567	10.16161	7.94771	9.83186
36	10.15354	11.67617	10.16161	11.67157	9.83186	11.56416
37	11.67617	7.92594	11.67157	7.93063	11.56416	8.03956
38	7.92594	9.86659	7.93063	9.85912	8.03956	11.88866

Table 3: For a queue with costs and parameters as given in Example 12 and discounted cost criteria with QoS measure L_1, the above table of iterations $\lambda_{i+1} = f(\lambda_i)$ shows that there are no equilibrium points but the iterates toggle between some intervals (generalized equilibrium sets defined in Section 3.7.2)



(a) (b)

Figure 24: For a queue with costs and parameters as given in Example 12 and discrete arrival distribution with discounted cost criteria: (a) QoS measure L_1 leads to neither an equilibrium set nor an equilibrium point but generalized equilibrium sets as defined in Section 3.7.2, (b) QoS measure L leads to an equilibrium point 13.41553.



(a) (b)

Figure 25: For a queue with costs and parameters as given in Example 12 and discounted cost criteria : (a) $f(\lambda, L_1)$ vs λ shows that there is no equilibrium point due to discontinuity; there exists generalized equilibrium sets (Section 3.7.2) (b) $f(\lambda, L)$ vs λ shows that the equilibrium point is 13.41553.

3.7.2 Generalized equilibrium sets and assumption $\tilde{A}1$

We first summarize some salient features of the above examples. These lead to relaxed notions of equilibrium sets and assumption A1 which we also discuss in detail.

In the last subsection, we saw that the definition of equilibrium set as given in Section 2.1 is not always valid in case of queues where arrivals have a particular form of discrete distribution because of the violation of assumption $A1$.

It was observed that there exist an equilibrium point when QoS measure L is used irrespective of the cost criteria. This might be attributed to the reason that L is a contraction mapping. This has been shown empirically in Table 4 where the contraction constant $l = \frac{|L(\lambda_2) - L(\lambda_1)|}{|\lambda_2 - \lambda_1|}$ is less than 1.

As seen in Example 9 and 10, even though there is an equilibrium point when L_1 is used, the iterates $\lambda_{n+1} = f(\lambda_n)$ do not converge as L_1 is not a contraction mapping. We show this using the data in Example 11. Consider $\lambda_1 = 10.6$ and $\lambda_2 = 10.625$ which belong to two different supports, then $L_1(\lambda_1) = 0.1983$, $L_1(\lambda_2) = 0.1719$, $f(\lambda_1) = 10.430$ and $f(\lambda_2) = 10.90$. Now, we have,

$$\frac{L_1(\lambda_1) - L_1(\lambda_2)}{\lambda_1 - \lambda_2} = 1.0531 > 1 \text{ and } \frac{f(\lambda_1) - f(\lambda_2)}{\lambda_1 - \lambda_2} = 18.9575 \; > 1$$

Therefore, L_1 and inturn $f(\lambda, L_1(\lambda))$ is not a contraction mapping.

We saw that in Example 7, there exist equilibrium sets consistent with the definition given in Section 2.1. These equilibrium sets are due to multiple optimal control limits which occur at change of support. Here, we observe that the overall effect of multiple optimal control limit and change in support makes the f locally monotone.

Apart from multiple optimal control limits, change of support can also lead to equilibrium set $E = E_1 \cup E_2$ but along with an invariant/equilibrium set E_0. This has been observed in Example 9 and 10 where change of support causes non monotonicity in $f(\cdot)$ and hence equilibrium sets E and E_0.

After observing the equilibrium behaviour in Example 11 and Example 12, we give a generalized definition of equilibrium set as follows:

Definition 2. ***(Generalized Equilibrium Set)*** *For a given $f(\lambda)$, if there exists a set $\tilde{E}$, $\tilde{R}$ and finite integers n_1 and n_2 such that*

- *for $\lambda \in \tilde{E}$, $f^{n_1}(\lambda) \in \tilde{E}$,*
- *for $\lambda \in \tilde{R}$, $f^{n_2}(\lambda) \in \tilde{E}$*

then $\tilde{E}$ and $\tilde{R}$ are called generalized equilibrium set and generalized repelling set respectively (f^n is the n^{th} composition of $f(\lambda)$).

In other words, a set $\tilde{E}$ is a generalized equilibrium set if the λ iterates visits this set infinitely often (when started in it, a subsequence of $\{\lambda_i\}_{i \geq 0}$ iterates visit it). These equilibrium sets could be interpreted as low, medium and high regime of arrival rates in the queue ($\tilde{E}_1$, $\tilde{E}_2$ and $\tilde{E}_3$ in Example 11 and Example 12.)

S. no	λ	**L (Ex 7)**	**l**	**L (Ex 8)**	**l**	**L (Ex 9)**	**l**	**L (Ex 10)**	**l**	**L (Ex 11)**	**l**	**L (Ex 12)**	**l**
1	3.31	0.00062	0.00390	0.00013	0.00129	0	0.00000	0	0.00000	0	0.00010	0.00361	0.01371
2	4.36	0.00472	0.01121	0.00149	0.00561	0	0.00005	0	0.00018	0.00010	0.00471	0.01801	0.02481
3	5.56	0.01817	0.01700	0.00822	0.01114	0.00006	0.00032	0.00022	0.00022	0.00575	0.00847	0.04777	0.02781
4	6.46	0.03346	0.01797	0.01825	0.01412	0.00035	0.00072	0.00042	0.00074	0.01338	0.01141	0.07281	0.02282
5	7.06	0.04424	0.01680	0.02672	0.01479	0.00078	0.00106	0.00087	0.00089	0.02022	0.01254	0.08650	0.01794
6	7.66	0.05432	0.01539	0.03560	0.01466	0.00142	0.00140	0.00140	0.00046	0.02775	0.01294	0.09727	0.05189
7	8.41	0.06587	0.00888	0.04659	0.01070	0.00247	0.00277	0.00106	0.00014	0.03745	0.04553	0.05835	0.00041
8	8.86	0.06987	0.00575	0.05140	0.00819	0.00122	0.00043	0.00112	0.00001	0.01697	0.00423	0.05854	0.00229
9	9.31	0.07245	0.00280	0.05509	0.00168	0.00142	0.00002	0.00111	0.00047	0.01887	0.00131	0.05751	0.01169
10	9.61	0.07161	0.00528	0.05560	0.00073	0.00142	0.00012	0.00097	0.00048	0.01927	0.00013	0.05400	0.01272
11	9.91	0.07003	0.00753	0.05538	0.00300	0.00139	0.00025	0.00083	0.00170	0.01931	0.00099	0.05018	0.01342
12	10.21	0.06777	0.00616	0.05448	0.00740	0.00131	0.00247	0.00032	0.00208	0.01901	0.00335	0.04615	0.08451
13	10.81	0.06407	0.00843	0.05004	0.00813	0.00279	0.00095	0.00157	0.00054	0.01700	0.00396	0.09686	0.00917
14	11.11	0.06660	0.00706	0.05248	0.00703	0.00308	0.00086	0.00173	0.00046	0.01819	0.00350	0.09961	0.00721
15	11.41	0.06872	0.00776	0.05458	0.00760	0.00333	0.00098	0.00187	0.00055	0.01924	0.00385	0.10177	0.00826
16	11.86	0.07221	0.00736	0.05801	0.00728	0.00377	0.00100	0.00211	0.00056	0.02097	0.00378	0.10549	0.00772
17	12.46	0.07663	0.00691	0.06237	0.00690	0.00437	0.00102	0.00245	0.00057	0.02324	0.00370	0.11013	0.00713
18	13.21	0.08181	0.00652	0.06755	0.00656	0.00514	0.00103	0.00288	0.00058	0.02601	0.00360	0.11548	0.00664
19	13.66	0.08475	0.00621	0.07050	0.00628	0.00560	0.00103	0.00314	0.00058	0.02763	0.00351	0.11847	0.00626
20	14.26	0.08847	0.00587	0.07426	0.00597	0.00622	0.00103	0.00349	0.00057	0.02974	0.00340	0.12222	0.00586
21	14.86	0.09199	0.00559	0.07785	0.00571	0.00684	0.00103	0.00383	0.00057	0.03178	0.00331	0.12574	0.00554
22	15.31	0.09451	0.00532	0.08042	0.00546	0.00731	0.00102	0.00409	0.00056	0.03327	0.00320	0.12823	0.00523
23	15.91	0.09770	-	0.08369	-	0.00792	-	0.00442	-	0.03519	-	0.13137	-

Table 4: For queues with discrete support arrival rate distribution, $L(\lambda)$ is seen to be a contraction mapping as $l = \frac{|L(\lambda_2)-L(\lambda_1)|}{|\lambda_2-\lambda_1|} < 1$ for the examples considered earlier. The values of λ have been chosen from different supports.

Remark 7. *The sets E_1 and E_2 from Definition 1 are special cases of generalized equilibrium sets. In fact, if $\lambda \in E_1$, then $f^{2n+1} \in E_2$ and $f^{2n+2} \in E_1$. These E_1 and E_2 are singletons for $M/M/1$, $D/M/1$ and some cases of discrete support queues also (Example 9 and Example 10).*

Remark 8. *The generalized repelling set $\tilde{R}$ can be equal to the complement of the union of generalized equilibrium sets ($\tilde{R} = (0, m) \setminus \tilde{E}$) as in Example 11 and Example 12. However, this might not always be true as in Example 9 and Example 10 where there is no repelling set; instead another equilibrium/invariant set E_0 exists.*

Remark 9. *Equilibrium point (λ_{eq}) is a special case of generalized equilibrium set in the sense that we can identify an ϵ neighbourhood around λ_{eq} such that the Definition 2 is satisfied.*

Remark 10. *However, we observe in Example 9 and 10 that even though the equilibrium point is inside generalized equilibrium set E_0, the iterates do not converge to it.*

Remark 11. *Also, there are examples where there is an equilibrium point and generalized equilibrium set not in the ϵ neighbourhood of the equilibrium point. We observe this situation in Example 11. We also observe that in Example 11, the ϵ neighbourhood around the equilibrium point at $\lambda = 8.3998$ acts a repelling set and the iterates visits $\tilde{E}_1$, $\tilde{E}_2$ and $\tilde{E}_3$ infinitely often.*

A relaxation of assumption $A1$ to hold in the case of queues with discrete support for arrival rate could be as follows :

Assumption 1′ ($\tilde{\mathbf{A}}\mathbf{1}$): For a given function $f(\lambda, L(\lambda, R^*_\lambda)) : (0, m) \to (0, m)$, there exists a monotonically decreasing function $\tilde{f}(\lambda, L(\lambda, R^*_\lambda)) : (0, m) \to (0, m)$ such that,

1. $\underline{f}(\cdot, \cdot) \leq f(\cdot, \cdot) \leq \bar{f}(\cdot, \cdot)$ and $\underline{f}(\cdot, \cdot) \leq \tilde{f}(\cdot, \cdot) \leq \bar{f}(\cdot, \cdot)$ where $\underline{f}(\cdot, \cdot)$ and $\bar{f}(\cdot, \cdot)$ are point wise bounds (envelopes) of $f(\cdot, \cdot)$ and

2. Area under $f(\cdot, \cdot)$ is same as area under $\tilde{f}(\cdot, \cdot)$, i.e.,

$$\int_0^m f(\lambda, L(\lambda, R^*_\lambda))d\lambda = \int_0^m \tilde{f}(\lambda, L(\lambda, R^*_\lambda))d\lambda \tag{9}$$

Under $\tilde{A}1$ we still have a non increasing function $\tilde{f}(\cdot, \cdot)$ that is sandwiched between the upper and lower bounds of $f(\cdot, \cdot)$; we pick the one such that the area under it is same as that under the $f(\cdot, \cdot)$ curve.

We illustrate the validity of this assumption for Example 12 in Figure 26 and observe that $\tilde{f}(\lambda)$ is indeed monotone.

4 Discussion

We demonstrated equilibrium sets in the interaction of queues with their service level sensitive customer bases. Equilibrium sets capture the oscillations between the high and low rate arrival regimes

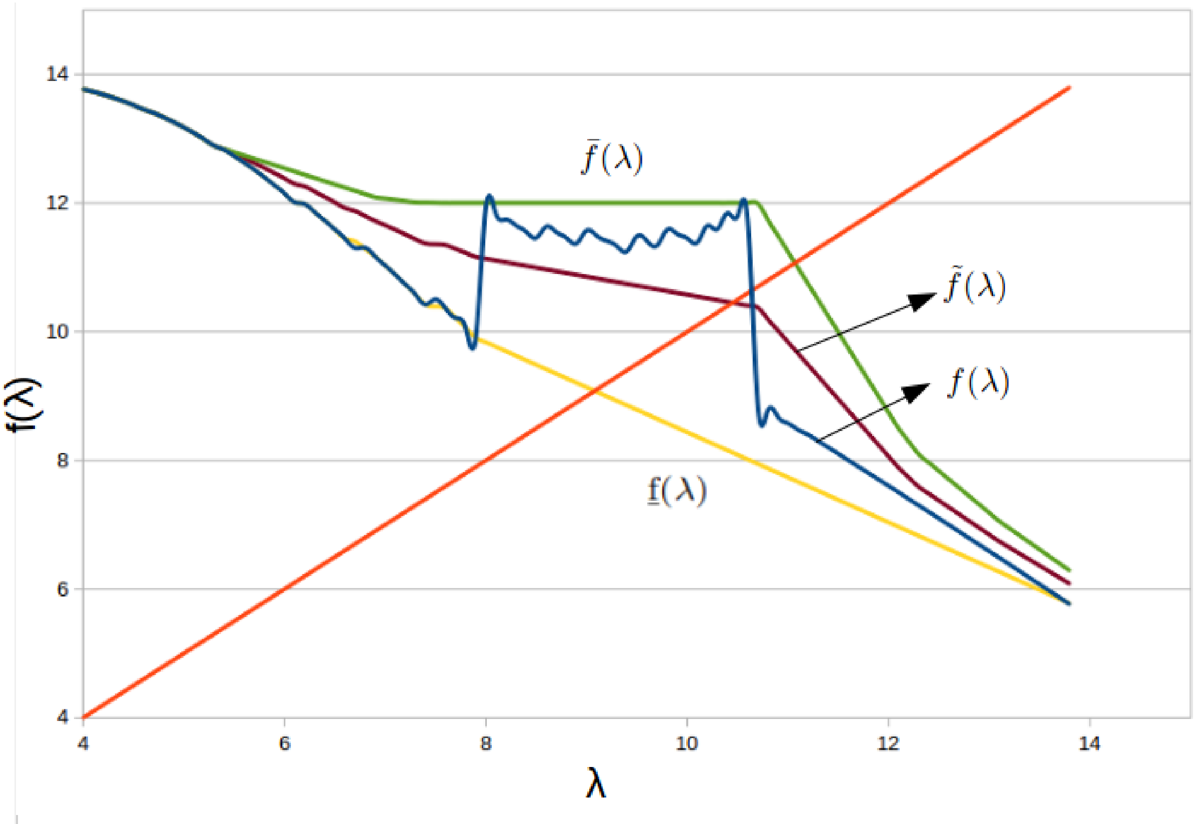


Figure 26: For Example 12, we show that $\tilde{f}(\lambda)$ is a monotone function and is sandwiched between $\bar{f}(\lambda)$ and $\underline{f}(\lambda)$. The area under $f(\lambda)$ and $\tilde{f}(\lambda)$ is approximately equal to 101.5 sq. units.

that one observes in the operation of queues. Some of the factors that can be attributed to this phenomena include multiple revenue optimal policies for the queue as shown for the stylized models of queues, the $M/M/1$ and $D/M/1$ queues. For the $M/M/1$ queues, the support is whole of R_+ and for $D/M/1$ queues the support changes everywhere. So we considered a particular $GI/M/1$ queue with an arrival rate that is continuous over its discrete support and change in support of their arrival distribution leads to more interesting equilibrium behaviour. Assumption A1 may not hold for such queues. However, oscillations between more than two arrival rate regimes can still be noted (the arrival rate iterates lie in some sets infinitely often) leading to the definition of generalized equilibrium sets. Thus, the equilibrium behaviour is more complicated in the absence of A1. So, more investigation of the existence as well as the characterization of such equilibrium sets in above setting for queues in general is promising.

Our results are likely to hold for multi-server variants of the queues we considered as well as with more general arrival streams since the results of (3) and (6) hold for such queues. A more interesting case would, of course, be the networks of queues.

It is known that strategic interactions usually lead to non-Pareto outcomes; for example, notions like Price of Anarchy (20) (a more suitable name would be 'price of free will' as in Stidham (19)) quantify the consequences. A potential topic would then be to investigate loss in *efficiency* of the above system comprising of a queue and its user-set with a suitably defined notion of efficiency.

A different, but, related theme would be to devise provable, and perhaps scalable, schemes to *predict* equilibrium sets from suitable data of a queue interacting with its user-set.

A Existence of control limits and properties of QoS

A.1 Proof of Theorem 1

Proof. Let R_λ is control limit when the arrival rate is λ of admission controlled $M/M/1$ queue. Let Q_{R_λ} denotes rate at which customers are admitted into the system in long run, H_{R_λ} denotes average queue length and $T(r, h, R_\lambda) = rQ_{R_\lambda} - hH_{R_\lambda}$ denotes the ergodic cost. Using the stationary distribution for $M/M/1/R_\lambda$ queue,

$$Q_{R_\lambda} = \frac{\lambda(1-\rho^{R_\lambda})}{1-\rho^{R_\lambda+1}} \text{ and}$$

$$H_{R_\lambda} = \frac{\rho(1-\rho^{R_\lambda})}{(1-\rho)(1-\rho^{R_\lambda+1})} - \frac{R_\lambda \rho^{R_\lambda+1}}{(1-\rho^{R_\lambda+1})}$$

when $\rho = \frac{\lambda}{\mu} < 1$. A finite control limit exists, if there exists R_λ such that, $T(r, h, R_\lambda - 1)$ is greater than $T(r, h, R_\lambda)$. Now, consider

$$T(r, h, R_\lambda) - T(r, h, R_\lambda - 1) = r\left(Q_{R_\lambda} - Q_{R_\lambda - 1}\right) - h\left(H_{R_\lambda} - H_{R_\lambda - 1}\right) \tag{10}$$

Putting the values of Q_l and H_l in equation (10) and simplifying, we get

$$T(r, h, R_\lambda) - T(r, h, R_\lambda - 1) = r\lambda\rho^{R_\lambda - 1}(1-\rho)^2 - h\rho^{R_\lambda}(-\rho - R_\lambda\rho + R_\lambda + \rho^{R_\lambda+1}) \tag{11}$$

From (11) and since $\rho < 1$, we conclude that one can find a sufficiently large R_λ for which $T(r, h, R_\lambda) - T(r, h, R_\lambda - 1) < 0$.

□

A.2 Proof of Theorem 4

Proof. Let $\{\tau_i\}_{i\geq 1}$ be the sequence of inter-arrival times and y be expected inter-arrival time. Also, let $\{\sigma_i\}_{i\geq 0}$ be the sequence of times of successive decision (arrival) epochs and W_t denote the state of the natural process (number in the system) at time t.

Let X_i, Y_i be the random variables denoting state and action at i^{th} decision epoch, $k(X_i, Y_i)$ be the lump sum portion of the reward and $c(W_t, X_i, Y_i)$ be the rate at which continuous portion of the reward is received between the decision epoch i and $i+1$ for $i = 0, 1, 2, \ldots$. From Chapter 11 of (5), the average reward can be defined as follows:

$$\begin{aligned} g^{R_\lambda}(s) &= \liminf_{n\to\infty} \frac{E_s^{R_\lambda}\left\{\sum_{i=0}^{n}\left[k(X_i, Y_i) + \int_{\sigma_i}^{\sigma_{i+1}} c(W_t, X_i, Y_i)dt\right]\right\}}{E_s^{R_\lambda}\left\{\sum_{i=0}^{n}\tau_i\right\}} \\ &= \liminf_{n\to\infty} \frac{\text{Expected total reward upto } n^{th}\text{decision epoch}}{\text{Expected total time until the } n^{th}\text{decision epoch}} \end{aligned}$$

Substituting $r = 1$ and $h = 0$ and noting that this is a finite state Markov chain, we have,

$$\begin{aligned} g^{R_\lambda}(s) &= \lim_{n\to\infty} \frac{E_s^{R_\lambda}\left\{\sum\limits_{i=0}^{n} k(X_i, Y_i)\right\}}{E_s^{R_\lambda}\left\{\sum\limits_{i=0}^{n} \tau_i\right\}} \\ &= \lim_{n\to\infty} \frac{\text{Expected number of customers admitted upto } n^{th} \text{ decision epoch}}{\text{Expected number of arrivals upto } n^{th}\text{decision epoch} \times E(\tau_1)} \\ &= \lim_{n\to\infty} \frac{\text{Expected number of customers admitted upto } n^{th}\text{decision epoch}}{n \times y} \end{aligned}$$

Now, consider

$$\begin{aligned} g^{R_\lambda}(s) \times y &= \lim_{n\to\infty} \frac{\text{Expected number of customers admitted upto } n^{th}\text{decision epoch}}{n} \\ &= \text{Fraction of customers admitted in the long run} \end{aligned}$$

One can compute g^{R_λ} by substituting $r = 1$ and $h = 0$ in Equation 13 in (6) after making the changes given in Section 3.2. This is same as the system of equation in (8) and let the solution of this system of equation be $\tilde{g}^{R_\lambda}$. Then, we get that the fraction of customers lost in long run, $L(\lambda, R_\lambda)$= $(1-\tilde{g}^{R_\lambda}y)$ and rate of customers lost, $L_1(\lambda, R_\lambda) = \lambda(1 - \tilde{g}^{R_\lambda}y)$. □

B Multiple optimal control limits

B.1 Proof of Theorem 5

Proof. Let $R^*_{\lambda_0}$ be the unique control limit at λ_0. Then,

$$v_\alpha^{R^*_{\lambda_0}}(s) > v_\alpha^{R_{\lambda_0}}(s) \;\text{ where } R_{\lambda_0} \neq R^*_{\lambda_0} \;\; \forall\, s.$$

For the linear system of equations considered in Equation (5), the coefficient matrix w.r.t. the variables $v_\alpha^{R_\lambda}(0), \ldots, v_\alpha^{R_\lambda}(R_\lambda)$ is same as the matrix $(I - M_{R^*_{\lambda_0}})$ as given in Theorem 11.3.1 in (5) where $M_{R^*_{\lambda_0}}$ is the transition probability matrix corresponding the Markov Deterministic (MD) policy $R^*_{\lambda_0}$. Since the solution of this system of equations is the value vector corresponding to the policy $R^*_{\lambda_0}$ exists and is unique, the matrix $(I - M_{R^*_{\lambda_0}})$ is invertible. Due to the continuity of matrix inverse, the system of equations in (5) will have a solution $\phi_i(\lambda); i = 0, \ldots, R_\lambda$, which is continuous in a small enough neighbourhood (a, b) of λ_0 (can also be obtained using implicit function theorem (22), (2) on the above linear system), given below:

$$\begin{aligned} v_\alpha(0) &= \phi_0(\lambda) \\ v_\alpha(1) &= \phi_1(\lambda) \\ &\vdots \\ v_\alpha(R_\lambda) &= \phi_{R_\lambda}(\lambda) \end{aligned}$$

So, for a small enough interval (a, b) around λ_0 we have

$$v_\alpha^{R^*_{\lambda_0}}(s) > v_\alpha^{R_\lambda}(s) \ \ \forall \ \ s, \ \ \forall \ \ \lambda \in (a, b).$$

Therefore, in the interval (a, b) the optimal control limit will be $R^*_{\lambda_0}$. □

B.2 Proof of Theorem 6

Proof. From Theorem 5, the optimal control limit at λ_2 cannot be unique as it will give rise to an open interval around λ_2 overlapping with (λ_1, λ_2) on the left (and (λ_2, λ_3) on the right) leading to a contradiction on uniqueness of R^*_λ (and $\tilde{R}^*_\lambda$). So, multiple optima exists at the common point λ_2 of these open intervals. From Corollary 3 in (6), the value $v_\alpha^{R^*_\lambda}(s)$ is unimodal with respect to R^*_λ with possible ties at multiple optimal control limit which necessarily occur consecutively. Also

$$v_\alpha^{R^*_\lambda}(s) \geq v_\alpha^{\tilde{R}^*_\lambda}(s) \ \text{ for } \ \lambda < \lambda_2$$

Taking the limit $\lambda \uparrow \lambda_2$ on both sides and using the fact that continuous and unique functions have a unique limit, we get

$$\begin{aligned} \lim_{\lambda \uparrow \lambda_2} v_\alpha^{R^*_\lambda}(s) &\geq \lim_{\lambda \uparrow \lambda_2} v_\alpha^{\tilde{R}^*_\lambda}(s) \\ v_\alpha^{R^*_{\lambda_2}}(s) &\geq v_\alpha^{\tilde{R}^*_{\lambda_2}}(s) \end{aligned} \tag{12}$$

Also

$$v_\alpha^{\tilde{R}^*_\lambda}(s) \geq v_\alpha^{R^*_\lambda}(s) \ \text{ for } \ \lambda > \lambda_2$$

Taking $\lambda \downarrow \lambda_2$, we get

$$\begin{aligned} \lim_{\lambda \downarrow \lambda_2} v_\alpha^{\tilde{R}^*_\lambda}(s) &\geq \lim_{\lambda \downarrow \lambda_2} v_\alpha^{R^*_\lambda}(s) \\ v_\alpha^{\tilde{R}^*_{\lambda_2}}(s) &\geq v_\alpha^{R^*_{\lambda_2}}(s) \end{aligned} \tag{13}$$

From (12) and (13), we get

$$v_\alpha^{\tilde{R}^*_{\lambda_2}}(s) \ = \ v_\alpha^{R^*_{\lambda_2}}(s) \ \ \forall \ \ s$$

Suppose $\hat{R}^*_{\lambda_2}$ is optimal at λ_2 but $R^*_{\lambda_2}$ is not optimal at λ_2. Then,

$$v_\alpha^{\hat{R}^*_{\lambda_2}}(s) > v_\alpha^{R^*_{\lambda_2}}(s) \ \ \forall \ \ s$$

We also know that, $v_\alpha^{R^*_{\lambda_2}}(s)$ is a continuous function of λ_2. Therefore, we can always find an ϵ, small enough such that

$$v_\alpha^{\hat{R}^*_{(\lambda_2-\epsilon)}}(s) > v_\alpha^{R^*_{(\lambda_2-\epsilon)}}(s) \ \ \forall \ \ s \tag{14}$$

However, (14) contradicts the definition of R^*_λ as unique optimal control limit in (λ_1, λ_2). Hence,

$$v_\alpha^{\hat{R}^*_{\lambda_2}}(s) = v_\alpha^{R^*_{\lambda_2}}(s) = v_\alpha^{\tilde{R}^*_{\lambda_2}}(s) \ \ \forall \ \ s \tag{15}$$

Since, $\hat{R}^*_\lambda$ is optimal control limit and (15) holds, it can be concluded that R^*_λ, $\tilde{R}^*_\lambda$ and $\hat{R}^*_\lambda$ are optimal at λ_2 and others also if they exist. □

B.3 Proof of Theorem 7

Proof. If there are multiple optima for a given λ, then,

$$v_\alpha^{R_\lambda}(s) = v_\alpha^{R_\lambda+1}(s) \quad \forall\, s \in S$$

The value vector can be uniquely determined by the first $R_\lambda + 1$ states, as the remaining $v_\alpha^{R_\lambda}(s)$'s for $s > R_\lambda$ can be evaluated recursively using the first $R_\lambda + 1$ components of the value vector. Therefore, it is enough to check that the above equation holds for all $s \le R_\lambda$.

As given by Van Nunen and Puterman in (3)

$$\begin{aligned} v_\alpha^{R_\lambda}(s) &= a_{s-1} v_\alpha^{R_\lambda}(s-1) - b_s \\ v_\alpha^{R_\lambda+1}(s) &= a_{s-1} v_\alpha^{R_\lambda+1}(s-1) - b_s. \end{aligned}$$

Now, $v_\alpha^{R_\lambda}(s) = v_\alpha^{R_\lambda+1}(s)$ gives $v_\alpha^{R_\lambda}(s-1) = v_\alpha^{R_\lambda+1}(s-1)$ and doing this recursively ultimately leads to $v_\alpha^{R_\lambda}(0) = v_\alpha^{R_\lambda+1}(0)$. So, this condition is a necessary one. We can see that it is also a sufficient condition.

□

B.4 Methods for calculating λ when $R_\lambda^* = 1, 2$ in a discounted cost criteria $M/M/1$ queue

After showing the existence of multiple control limits, we want to identify the point λ where multiple optimal control limits exist. Following three methods involve finding $\Delta\lambda$, if we start from some arbitrary λ_0. We only consider the case when $R_\lambda = 1, 2$. For higher value of control limits, these methods becomes very cumbersome.

B.4.1 Method 1

The first method involves equating $v_\alpha^{R_\lambda}(0)$ and $v_\alpha^{R_\lambda+1}(0)$ as a function of $\Delta\lambda$. In particular, for $R_\lambda = 1$ and $R_\lambda = 2$, value at state 0 is given as folllows:

$$\begin{aligned} v_\alpha^1(0) &= \frac{r(1-p_0) - c(1)}{1 - q_1 - p_0} \\ v_\alpha^2(0) &= \frac{r(1 - p_1 - p_0^2) - c(1)(1 - p_0 - p_1) - p_0 c(2)}{(1-q_1)(1 - p_0 - p_1) - p_0 q_2} \end{aligned}$$

After adding each of the roots obtained from solving $v_\alpha^1(0) = v_\alpha^2(0)$ to λ_0, one of these value is the λ where the multiple optimal control exists. However, we need to check and find which one is the correct root by finding the control limit at the each of them. Also, some roots might be invalid. Following example illustrate the method for a $M/M/1$ queue with given set of costs and parameters r, h, μ, α. The computations were done in Mathematica 9 on Paaspoli server (OS : Debian , AMD OPTERON 6212, 2.6 GHz processors and 32GB RAM).

1. $h = 1$, $r = 0.4$, $\mu = 5.5$, $\alpha = 0.54$ and $\lambda_0 = 10$
   The simplified version of the final equation to be solved is $1.45185\Delta\lambda^4 + 61.9742\Delta\lambda^3 + 955.654\Delta\lambda^2 + 6309.78\Delta\lambda + 14988.1$. Roots for this equation are

   (a) $\Delta\lambda_1 = -10.000000$

   (b) $\Delta\lambda_2 = -6.106282$

   (c) $\Delta\lambda_3 = -10.540051$

   (d) $\Delta\lambda_4 = -16.040032$

   Root number (b) is the point which after adding to λ_0 gives the point of multiple optima i.e. $\lambda = 3.893718$.

B.4.2 Method 2

This method involves solving an optimization problem which minimizes the square of the difference between the value corresponding to control limits $R_\lambda = 1, 2$ at state zero. The optimization problem is formulated as follows:

$$M2 \qquad \min_{\Delta\lambda} \; (v_\alpha^{1,0}(\lambda_0 + \Delta\lambda) - v_\alpha^{2,0}(\lambda_0 + \Delta\lambda))^2$$

subject to

$$\begin{aligned} v_\alpha^{1,0}(\lambda_0 + \Delta\lambda) &= \frac{r(1-p_0) - c(1)}{(1-p_0-p_1)} \\ v_\alpha^{2,0}(\lambda_0 + \Delta\lambda) &= \frac{r(1-p_0^2-p_1) - (1-p_0-p_1)c(1) - p_0c(2)}{(1-q_1)(1-p_1-p_0) - p_0q_2} \end{aligned} \tag{16}$$

$\Delta\lambda$ unrestricted

The above constrained optimization problem is a non convex minimization problem. It can be solved easily using solvers like SNOPT, KNITRO etc. This approach has an edge over the earlier one as it gives an if and only if condition for obtaining the λ where multiple optimal control limits exist.

Theorem 10. *For a $M/M/1$ queue with given costs and parameters, multiple optimal control limits 1 and 2 will exist if and only if $M2$ attains its global minima.*

Proof. Suppose we have multiple optimal control limits 1 and 2 for a given λ_m (say $\lambda_0 + \Delta\lambda_m$). Then, at $\lambda_0 + \Delta\lambda_m$, the value vectors for control limit 1 and 2 are equal i.e.

$$v_\alpha^{1,s}(\lambda_0 + \Delta\lambda_m) = v_\alpha^{2,s}(\lambda_0 + \Delta\lambda_m) \;\; \forall\, s$$

In particular, we have $v_\alpha^{1,0}(\lambda_0 + \Delta\lambda_m) = v_\alpha^{2,0}(\lambda_0 + \Delta\lambda_m)$. Hence, we conclude that at the point of multiple optima λ_m, the objective value of $M2$ is 0 and it attains its global minima. Let us consider the converse case where we are given that the global optimal solution of $M2$ with objective value 0 is attained at $\Delta\lambda_m$. Then, we have that for $\Delta\lambda_m$,

$$v_\alpha^{1,0}(\lambda_0 + \Delta\lambda_m) = v_\alpha^{2,0}(\lambda_0 + \Delta\lambda_m)$$

This is a sufficient condition for multiple optimal control limit to exist at $\lambda_0 + \Delta\lambda_m$ as given in Theorem 7.

Hence, the solution set of the above minimization problem characterizes the set of points at which multiple optimal control limit exists. $\square$

Following example illustrate this constrained optimization method. The optimization problem was modeled in AMPL and solved using SNOPT solver on OPTIMUS server (OS Linux, Intel Quad core Xeon E5506 2.13 GHz and 64GB RAM).

1. Costs and parameters $h = 1$, $r = 0.4$, $\mu = 5.5$, $\alpha = 0.54$ and $\lambda_0 = 5$ of $M/M/1$ queue when used in $M2$ gives, objective value = $1.349202e^{-12}$ and $\Delta\lambda = -1.10629$ giving $\lambda = 3.89371$.

B.4.3 Method 3

If we substitute the value of $v_\alpha^{1,0}(\lambda_0 + \Delta\lambda)$ and $v_\alpha^{2,0}(\lambda_0 + \Delta\lambda)$ in objective function of (16), then we get the unconstrained version of $M2$ given as follows :

$$M3 \qquad \min_{\Delta\lambda} \left(\frac{r(1-p_0) - c(1)}{(1-p_0-p_1)} - \frac{r(1-p_0^2-p_1) - (1-p_0-p_1)c(1) - p_0 c(2)}{(1-q_1)(1-p_1-p_0) - p_0 q_2} \right)^2 \tag{17}$$

$$\Delta\lambda \ \ \text{unrestricted}$$

It can be solved by equating the gradient of the objective function in $M3$ to 0 as shown below.

$$2(v_\alpha^1(\Delta\lambda) - v_\alpha^2(\Delta\lambda)) \frac{d(v_\alpha^1(\Delta\lambda) - v_\alpha^2(\Delta\lambda))}{d\lambda} = 0 \tag{18}$$

The solution of unconstrained optimization problem can be root of either of the terms in (18). However, we are only interested in the root of the first term which gives global minima. The roots of second term might lead to some unwanted solution with non-zero objective value.

Following example illustrate this method using AMPL as modelling language. The optimization problem was solved using SNOPT solver on OPTIMUS server (OS Linux, Intel Quad core Xeon E5506 2.13 GHz and 64GB RAM).

1. Costs and parameters $h = 1$, $r = 0.4$, $\mu = 5.5$, $\alpha = 0.54$ and $\lambda_0 = 10$ of $M/M/1$ queue when used in $M3$ gives, objective value = $3.722129e^{-10}$ and $\Delta\lambda = -6.10629$ giving $\lambda = 3.89371$.

C Proof of Theorem 8

Proof. First consider a queue with discounted cost criteria and QoS measure L. For a $M/M/1$ queue, the entries of the transition probability matrix of the embedded Markov chain are continuous functions of rate λ. This gives us that the value obtained from maximization of Equation (4) is a continuous function of λ and the set of optimal policies is a continuous set. Using this, we showed in Section 2.1 that QoS measures L and L_1 are piecewise continuous over $(0, m)$ (9).

We are given that there exists an equilibrium set for a $M/M/1$ queue. This implies that there exists a point $\lambda = l_0$ where $f(\lambda)$ is discontinuous. Because, if the optimal control limit at l_0 is unique, then from Theorem 5 there exists an open interval $(l_0 - \epsilon, l_0 + \epsilon)$ over which the control limit is R^*_λ and $f(\cdot)$ can not be discontinuous at l_0.

Further, using the same argument as in proof of Theorem 6, we can identify optimal control limits $R^*_{2,\lambda}$ for $\lambda \in (l_0 - \epsilon, l_0)$ and $R^*_{1,\lambda}$ for $\lambda \in (l_0, l_0 + \epsilon)$ such that at l_0 both $R^*_{2,\lambda}$ and $R^*_{1,\lambda}$ are optimal. Thus a discontinuity in $f(\cdot)$ at l_0 due to such multiple optimal control limits leads to the existence of equilibrium sets. The above arguments are valid for other combinations of cost criteria and QoS measures. $\square$

D Generation of the arrival rates from a discrete distribution

Consider the following discrete distribution whose mean is *required* to be locally continuous:

$$G(.) = \begin{cases} u_{1,k} & \text{when arrival rate is } 1 \\ u_k & \text{when arrival rate is } k \\ u_{k+1} & \text{when arrival rate is } k+1 \end{cases}$$

Mean arrival rate $\lambda = u_{1,k} + k \times u_k + (k+1) \times u_{k+1}$. We will fix $u_{1,k}$ in order to generate mean arrival rate for a suitable interval. Let us take the support as $\{1,\ k,\ k+1\}$; the arrival rate is $\lambda \in (\underline{\mathrm{m}}_k, \overline{m}_k]$ where, $\overline{m}_k$ and $\underline{\mathrm{m}}_k$ can be derived as follows,

$$\begin{aligned} \lambda &= u_{1,k} + k \times u_k + (k+1)u_{k+1} \\ &= u_{1,k} + k \times u_k + (k+1)(1 - u_{1,k} - u_k) \\ &= (k+1) - k \times u_{1,k} - u_k \end{aligned}$$

Now, $u_k \in [0,\ 1 - u_{1,k}]$. Thus, $\overline{m}_k = (k+1) - k \times u_{1,k}$ and $\underline{\mathrm{m}}_k = (k+1) - k \times u_{1,k} - (1 - u_{1,k})$. To generate arrival rate whose mean is $\lambda \in (\underline{\mathrm{m}}_{k+1}, \overline{m}_{k+1}]$, we change the support of the distribution to $\{1,\ k+1,\ k+2\}$ and so on. While changing the support we are fixing $u_{1,k+1} \geq u_{1,k}$, where $\overline{m}_{k+1} = (k+2) - (k+1) \times u_{1,k+1}$ and $\underline{\mathrm{m}}_{k+1} = (k+2) - (k+1) \times u_{1,k+1} - (1 - u_{1,k+1})$.

Example 13. *In this example, we will illustrate the above method to generate arrival rates whose mean lies in* $(3.7,\ 6.25]$.

1. We take support as $\{1,\ 4,\ 5\}$ and fix $u_{1,4} = 0.1$. Then, $\lambda = 0.1 + 4u_4 + 5(0.9 - u_4)$, i.e., $\lambda = 4.6 - u_4$. Thus, with support $\{1,\ 4,\ 5\}$, $u_k \in [0, 0.9)$ and $u_5 = 1 - u_4$ we can generate a set of arrival rates with above mass functions such that the arrival rate λ is continuous over the interval $(3.7,\ 4.6]$.

2. Support now is $\{1,\ 5,\ 6\}$ and we fix $u_{1,5} = 0.15$. So, $\lambda = 0.15 + 4u_5 + 5(0.85 - u_4)$, i.e., $\lambda = 5.25 - u_5$. Thus, with $[1,\ 5,\ 6]$ support we can generate $\lambda \in (4.4,\ 5.25]$ but we have already generated λ till 4.6 from previous support; so, we generate a family of arrivals whose rates continuously change over $\lambda \in (4.6,\ 5.25]$.

3. We take support as $\{1,\ 6,\ 7\}$ and we take $u_{1,6} = 0.2$. Then, $\lambda = 0.2 + 6u_6 + 7(0.8 - u_6)$, i.e., $\lambda = 5.8 - u_6$. Thus, with support $\{1,\ 6,\ 7\}$ we can generate $\lambda \in (5,\ 5.8]$ but we have already generated λ till 5.25 from previous support; so, we now have arrivals whose rate is locally continuous over $\lambda \in (5.25,\ 5.8]$.

4. We take $\{1,\ 7,\ 8\}$ as support and $u_{1,7} = 0.25$ so that, $\lambda = 0.25 + 7u_7 + 8(0.75 - u_7)$, i.e., $\lambda = 6.25 - u_7$. Thus, with support $\{1,\ 7,\ 8\}$ we can generate $\lambda \in (5.5,\ 6.25]$ but we have already generated λ till 5.25 from previous support; so, we have the arrivals with the desired property over $\lambda \in (5.8,\ 6.25]$.

References

[1] Eren Basar Cil, E. Lerzan Ormeci, and Fikri Karaesmen. Effects of system parameters on the optimal policy structure in a class of queueing control problems, *Queueing Systems* 61, 273-304, 2009.

[2] Glenn James and Robert Clarke James. The Mathematics Dictionary, *Springer*, 1992.

[3] J. A. E. E. Van Nunen and Martin L. Puterman. Computing Optimal Control Limits for $GI/M/s$ Queuing Systems with Controlled Arrivals, *Management Science* 29(6), 725-734, 1983.

[4] Linn I. Sennott. Stochastic Dynamic Programming and the Control of Queueing Systems, *John Wiley & Sons*, 1998.

[5] Martin L. Puterman. Markov Decision Processes: Discrete Stochastic Dynamic Programming *John Wiley & Sons*, 2005.

[6] Martin L. Puterman and Lyn C. Thomas. A Note on Computing Optimal Control Limits for $GI/M/1$ Queueing Systems, *Management Science* 33(7), 939-943, 1987.

[7] Moshe Haviv and Martin L. Puterman. Bias optimality in controlled queueing systems, *J. Appl. Prob.*, 35(1), 136-150, 1998.

[8] N. Hemachandra, Sandhya Tripathi and Kishor Patil. Equilibrium sets of some $GI/M/1$ queues, *Communicated*.

[9] N. Hemachandra, Rajesh Kamma Sri Naga and Mohd. Abdul Qavi. A model for equilibrium in some service-provider user-set interactions, *Annals of Operations Research* 243, 95-115, 2016.

[10] Pinhas Naor. The Regulation of Queue Size by Levying Tolls, *Econometrica*, 37(1), 15-24, 1969.

[11] Rafael Hassin. Rational Queueing, *Chapman & Hall/ CRC Series in Operations Research, CRC Press*, 2016.

[12] Refael Hassin and Moshe Haviv. To Queue or Not to Queue, *International Series in Operations Research and Management Science, Springer Science & Business Media*, 2003.

[13] Ronald W. Wolff. Stochastic Modeling and the Theory of Queues, *Prentice-Hall International Series in Industrial and Systems Engineering*, 1989.

[14] Robert C. Rue and Matthew Rosenshine. Some properties of optimal control policies for entry to an $M/M/1$ queues, *Naval Research Logistics Quarterly*, 525-532, 1981.

[15] Shaler Stidham Jr. and Richard R. Weber. A survey of Markov decision models for control of networks of queues, *Queueing Systems* 13, 291-314, 1993.

[16] Shaler Stidham.Jr. Optimal Control of Admission to a Queueing System, *IEEE Transactions on Automatic Control* 30(8), 705-713, 1985.

[17] Shaler Stidham.Jr. Optimal Design of Queueing Systems, *Chapman & Hall CRC*, 2009.

[18] Shaler Stidham.Jr. Socially and Individually optimal control of arrivals to a $GI/M/1$ queue, *Management Science* 24, 1978.

[19] Shaler Stidham.Jr. The price of anarchy for a network of queues in heavy traffic, *Essays in Production, Project Planning and Scheduling: A Festschrift in Honor of Salah Elmaghraby* 91-121, 2014.

[20] Tim Roughgarden. Introduction to inefficiency in equilibria, *Chapter 17 in Algorithmic Game Theory, Cambridge University Press* 441-457, 2007.

[21] Whan S. Kim. Price-based quality-of-service control framework for two-class network services, *Journal of Communications and Networks* 9(3), 319329, 2007.

[22] Wendell H. Fleming. Functions of Several Variables, *Springer Science & Business Media*, 1987.